Print ISSN : 2395-1990
Online ISSN : 2394-4099
www.ijsrset.com

NEW HORIZON COLLEGE OF ENGINEERING
Autonomous College Permanently Affiliated to VTU, Approved by AICTE & UGC
Accredited by NAAC with 'A' Grade, Accredited by NBA

IJSRSET

TECHNO SCIENCE ACADEMY
THE INTERNATIONAL OPEN ACCESS PUBLISHER

Fourth National Conference on Advancements In Computer Science and Engineering

Organised by
Department of Computer Science and Engineering,
New Horizon College of Engineering
In Association with
VTU, Belagavi, Silver Jubilee Celebrations and
Computer Society of India

VOLUME 9, ISSUE 11, MAY-2022

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Email : editor@ijsrset.com Website : http://ijsrset.com

Fourth National Conference on Advancements in Computer Science and Engineering

NCACSE-2022

11th May, 2022

Organised by

NEW HORIZON COLLEGE OF ENGINEERING

Autonomous College Permanently Affiliated to VTU, Approved by AICTE & UGC

Accredited by NAAC with 'A' Grade, Accredited by NBA

Department of Computer Science and Engineering,

New Horizon College of Engineering, Bengaluru, Karnataka, India

In Association with

VTU, Belagavi, Silver Jubilee Celebrations and Computer Society of India

Publication with

IJSRSET

International Journal of Scientific Research in Science, Engineering and Technology

Online ISSN : 2394-4099 | Print ISSN : 2395-1990

Volume 9, Issue 11, May-2022

Published By

TECHNO SCIENCE ACADEMY

THE INTERNATIONAL OPEN ACCESS PUBLISHER

website : www.technoscienceacademy.com

Chief Patron

Dr.Mohan Manghnani
Chairman
New Horizon Educational Institution, Bengaluru

Patron

Dr.Karisiddappa
Vice chancellor,
Visvesvaraya Technological University, Belagavi

Dr.A.S. Deshpande
Registrar
Visvesvaraya Technological University, Belagavi

Dr.Manjunatha
Principal
New Horizon College of Engineering,Bengaluru

Advisory Committee

Dr.R.J.Anandhi, Dean Academics & Professor /HOD-ISE, NHCE,Bengaluru
Dr.Annappa Basava, Professor, National Institute of Technology, Surathkal
Mr.S.A.Satish, Project Delivery Head, TCS, Bengaluru
Dr.Ajit Danti, Professor & Program Coordinator for B.Tech CSE, Christ University ,Bengaluru
Dr.P.Deepalakshmi, Professor, Dean & HOD, CSE, School of Computing, Kalasalingam University, Krishnan Kovil, TamilNadu

Conference Chair

Dr.B.Rajalakshmi, Professor & HOD-CSE, NHCE
Dr.Siddesh G K,Head R & D, NHCE
CONFERENCE CO – CHAIR
Dr.B.Meenakshi Sundaram, Professor/CSE, NHCE
Dr.S.J.Subhashini, Associate Professor/CSE, NHCE
Ms. Lakshmi S Hanne, Assistant Professor/CSE, NHCE
List of Committee Members
Dr. Amarjeet Singh
Dr. Rachana P
Dr.Senthil Kumar R
Dr.Nirmala M
Mr. Siva Balan N

Ms. Asha Rani Borah
Ms. Shanmuga Priya S
Ms. Pramilarani K
Ms. Soja Rani S
Mr. Santosh Kumar B
Ms. Suganya.R
Mr. Manjunatha Swamy C
Ms. Uma N
Ms. Subhashree Rath
Ms. C Shreenagamanjula Rani
Ms.N Sasikala
Ms.Devi Naveen
Ms.Jayashree
Ms.Srividhya G
Ms.D.Salangai Nayagi
Ms.Jeevitha R
Ms. Yogitha
Ms. Tinu N S
Mr.M.Manikanda Kumar
Mr.Mahantesh K M
Mr. Bhaskar S V
Ms. Chempavathy
Ms.Bangari Sindhuja
Ms.Pandey Surya Sampurnand
Ms.Revathi S
Ms.Nayana N Kumar
Ms.Manjula A K
Ms.Shruthi K C
Ms.Florance G
Mr.Geluvaraj B
Ms.Priya T V

About Conference

The conference is intended to provide a common forum for discussions about the recent trends & advancements of research in the area of Computer Science. This conference is a very good platform for academia, researchers and technologists to discuss and present recent advances and research outcomes.

We invited technical papers on, domains like Networks, Big data analytics, Machine learning, Data Science, IoT, Network security, Cloud computing, Data Mining, Blockchain, to name a few.

About College

New Horizon College of Engineering is an Autonomous college affiliated to Visvesvaraya Technological University (VTU), approved by the All India Council for Technical Education (AICTE) & University Grants Commission(UGC). It is accredited by NAAC with 'A' grade & National Board of Accreditation (NBA). It is one of the top Engineering Colleges in India as per NIRF rankings, ARIIA – 2020 and an ISO 9001:2008 certified Institution. New Horizon college of Engineering is located in the heart of the IT capital of India, Bangalore. The college campus is situated in the IT corridor of Bangalore surrounded by MNCs and IT giants such as Intel, Accenture, Capegemini, ARM, Symphony, Wipro, Nokia, JP Morgan and Cisco to name a few.

NHCE has a scenic and serene campus that provides an environment which is conducive for personal and intellectual growth. The infrastructure acts as a facilitator for the effective delivery of the curriculum. NHCE boasts of state -of-the -art facilities for its students. The students are given utmost encouragement in their areas of interest by providing hi-tech facilities backed by faculty support. The institute places highest priority on innovative programs of instructions that include both traditional class room theory and professional skills training. There is a strong impetus on overall personality development of the students with emphasis on soft skills. Students are supported through mentoring and counseling systems. The management offers scholarships to meritorious students. At NHCE, from the moment a student walks into the campus, he/she is well guided to know his/her strengths and choose an area of functional specialization. This enables students to concentrate their efforts and energies to gain the competitive edge. NHCE has a unique distinction of achieving 100% admissions in all its courses year after year.

NHCE has Centre of Excellence with reputed industries like Adobe, HPE, Vmware, Schneider Electric, SAP, Quest Global, CISCO.

CONTENTS

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Homophily-Birds of A Feather Flock Together

Shivangi, Shreeyam
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Similarities in racial relationships: This principle of homosexuality builds network relationships in all areas, including marriage, kinship, friendship, work. support or information transfer of all kinds, exchanges between nodes, fellowship membership, and other types of relationships. increase. As a result, people's personal networks are homogeneous in many demographics, including socio-demographics, behaviour, and individual personal characteristics. Therefore, it basically limits the social and personal world of people / actors in a way that strongly implements the information they receive, the attitudes and perspectives they form, and the interactions they experience between them. To do. Homosexuality creates a great division in our personal environment in terms of age, religion, education, profession, and gender when we look at race and ethnicity. Geographical injustices, family ties, organizations, and homogenous positions in social systems all create context in which homogenous relationships are formed. Bonds between heterogeneous people also break at a higher rate and set the stage for forming a local position in the social arena. Discuss for further research. (A) There are basic ecological processes that connect multiple organizations, associations, cultural and non-cultural communities, social and non-social movements, and many other forms. (B) The effect of multiple bonds on the homosexual matrix. (C) The dynamics of network modeling change over time, co-evolving networks and other social units.

Keywords: humanecology ; voluntary associations ; organizations .

I. INTRODUCTION

Individuals with different characteristics, such as gender, race, ethnicity, age, class background, and education level, seem to have very different traits. These differences are often attributed to some aspect of their category membership. The emotional nature of women, the tolerance of educated people, and the violent nature of gang members are among these qualities. Despite the significant differences in the worlds that these people inhabit, essentialists ignore these differences. Generally only people of the same socioeconomic group come into contact with each other, so any quality tends to become localized. If we interact with others who hold the same views as us, any experiences we have as a result of our position are reinforced. It comes to typify “people like us.” Homophily refers to the phenomenon of similar people contacting each other at a greater rate than dissimilar people. As a result of homophily,

information that flows through a network tends to be localized, whether that is cultural, behavioral, genetic, or material. In homophily, the distance in terms of social characteristics is translated into the number of relationships needed to connect two individuals, or the network distance. Furthermore, any social entity that has a substantial dependence on networks for its transmission tends to be localized in a social space and will facilitate certain dynamics in the manner in which it interacts with other social entities in an ecological system of social forms. Studies of networks include studies of social networks, voluntary associations, social capital (at the individual and community levels), social movements, culture, organizations, and a range of other topics that are influenced by networks. This principle is vital to the way in which they operate, so we use it for our organizing principle. We review the classical applications of the concept after summarizing the extensive empirical evidence for the theory. It has been shown that a wide range of aspects on which homophily can exist can occur in network relationships, and that many types of relationships can be homophilous. As a follow-up to our discussion of homophily's sources, we examine social structures that encourage propinquity and cognitive processes that facilitate communication amongpeople who have similar interests. We conclude by suggesting avenues for future research.

Studies of Homophily Across the Century: Methodological and Substantive Progressions

Homosexual Studies in the Century: Advances in Method and Content It is important that the relationships among homosexuals be consistent in the initial studies of homosexuality. coffee) is easily revealed to ethnographic observers in small social enterprises. Disclosure. I aim for that. A lot of early regional studies focused primarily on demographic characteristics such as age, gender, race/ethnicity, level of education (Bott 1929, Loomis 1946, etc.), intelligence, attitudes, and aspirations (Almack 1922, etc.). It's done. Basically, it is a property based exchange. In a time when race issues and dropouts dominated American politics, many scholars found themselves segregated from newly abolished schools, buses, and public places. This trend continued. Furthermore, when the researcher examined the correlation between different attitudinal biases and sub-dimensions, they found strong similarity between linked styles.

It is assumed by second culture that friends are a major influence on the behavior of people, especially of young people.

II. RACE AND ETHNICITY

Clearly, race and ethnicity are the largest disparities on US social networks today, and their impact on establishing networks in other ethnically diverse societies is very significant.In this area, the basic homosexuality created by groups of different sizes is combined with and the future of different positions of racial / ethnic groups in other dimensions (education, profession, income, religions etc. A common personal prejudice is the highly visible and often investigated network disruption. Homosexuality is strongly correlated with ethnicity.Labor relations (Lincoln & Miller 1979, Ibarra 1995), limited networks of discussion about a particular topic (Schneider et al. 1997), the mere fact that they appear together in public (Mayhew et al. 1995) or someone who "knows about something" (Lawrence 2000).This pattern is evident even in negative attachments such as damage and rape (South & Felson 1990, South & Messner 1986).limitednetworks of discussion about a A particular topic (Schneider et al. 1997), the mere fact that they appear together in public .Even the negative attachment of damage and rape follows this pattern

(South & Felson 1990, South & Messner 1986). Only 8% of adults with networks of size 2 and above have people of different races who "discuss important issues." This represents less than one-seventh of the non-uniformity observed when humans choose randomly. This is less than one-seventh of the non-uniformity observed when humans choose randomly. This population (Marsden1987). Also, it is much more likely to report that familiars are connected if they are of the same race (Louch2000).Of course, people often refer to spouses and other relatives as best friends, so strong race-based homosexuality enhances the homosexuality of trust. However, the level of racial heterogeneity is still only one-quarter of the possibilities, even if only those who do not mention relatives in the discussion network are considered (Marsden1987).Basic homogeneity within most opportunity structures (the focus of the public, SMSA, workplace, and other activities) leads to orders. In the same way that people refer to their spouses and other relatives as best friends, strong race-based homosexuality enhances the homosexuality of trust.I showed that. Interestingly, African-American-Anglo contact can be an exception to this pattern. This is because their marriage rates are not well explained by the population distribution.

III. RESULT AND DISCUSSION

The dynamics of homosexuality-based opinions that birds flock tendency that flocks together. The more similar our opinions are, the more likely we are to interact and converge with each other. This interaction is unlikely. Here, we will generate multiple clusters depending on the size of the boundary with multiple clusters. With the introduction of noise, BCM creates consensus, leaving a solution for explaining the emergence, maintenance, and polarization of opinion clusters. To test the relevance of homophily we distinguish Between Baseline homophily: simply the amount of homophily that we would expect that by chance given an uneven distribution of people with various characteristic. Inbreeding homophily: the amount of homophily over the above expected value is typically because of personal preferences and choices.

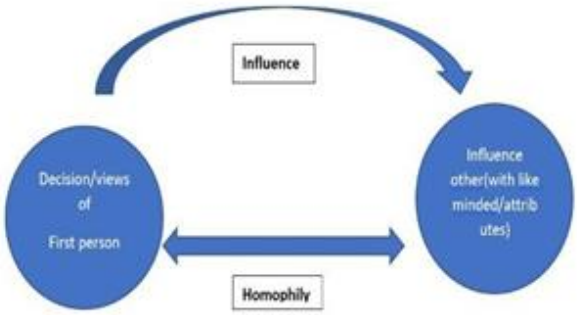


Fig. 1: Influence created by homophily

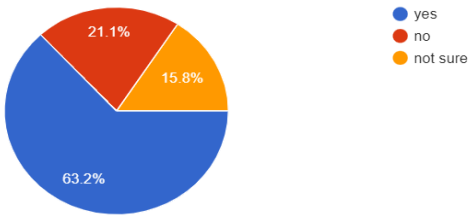
In social network analysis, instead ofanalyzing individual behaviors, attitudes and beliefs focus lies on actor/nodes.

for homophily as well this works for interaction with one another of same kind or of different what difference they make and how these interactions develope various scenario with one another.

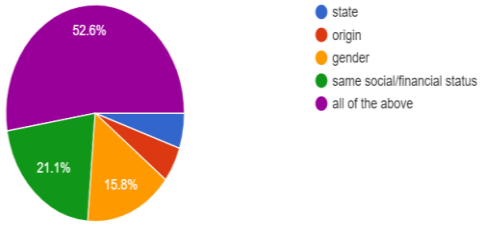
As per the above concept of wasserman and galaskiewiczand survey represent the influence of homophily in india.

We performed a survey on 20 different people in India and got the following results.

Do you think our Indian society is influenced with term homophily knowingly or unknowingly?

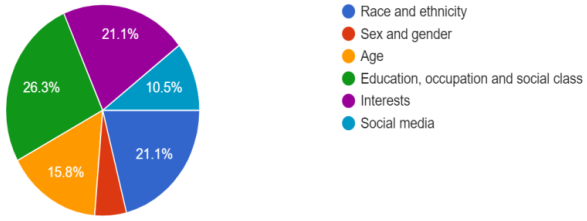


Mostly people are influenced based on various factor they are going through and living through as a origin. For what contributes more to homophily in India ?



opinions are based on state, origin, gender as well as same gender or financial status are chosen altogether. somewhere people tend to get influenced by something similar to them.

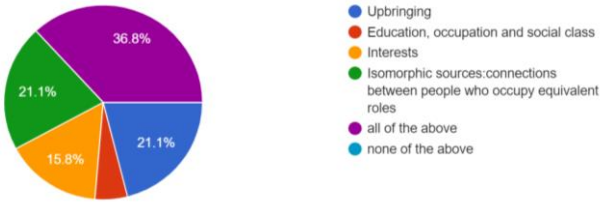
If we talk about domain based homophily we come along various domain influenced by homophily more or less.



We can see that race and ethnicity influences around 21.1% of homophily that tend us more toward homphily in terms of connecting to the people of same as they might think they will understand them what they are going through or they won't have to explain much to the person next to them. Age somehow less influenced as age grows gains maturity. Each individual consider as actor they tend to more influenced by education, occupation and social class.

As it depicts the influence of environment they are getting and openness to challenge to come forward to come over any restriction.

By this data we quench for more optimal reason to get influenced by homophily is that what make us more adaptive to homophily: upbringing, education environment, isomorphic sources, interest or all.



Is that all mentioned factors are influencing naturally to our mind and we are becoming more receptive to it.

IV. CONCLUSION

Conclusion to it based on various domain and factors is that homophily is omnipresent. each individual considered as nodes/actor. Tends to be friend with each other based on similarity based various domain discussed above.

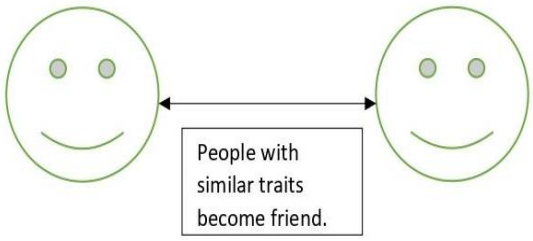


Fig. 3: Similar traits become friend

Here we discuss about conceptual principal of homophily based on design principles of the Network-Oriented Modeling. In this A,B,C,D are actors represented as xa, xb, xc, xd and ωA,B,(B,C),(A,D),(D,C) represents the weight of the nodes connected to them, which describes the features we can relate with each actors in term to seek for homophily nature among each individual actor.

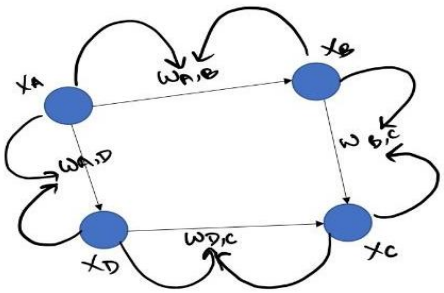


Fig. 4: Conceptual representation of homophily

The results shown bring initiative for new change is that now people will be aware of their nature of making bond/network among same kind of group/tribe or people. Although we cannot model the exact the imprint of network formation.But we can establish a mechanism through this analysis through these data that same kind of actors form a ties among them.regardless of weak ties and strong ties they form local bridge and bridge as well. It appears among large community of people through various domain and presence of awareness is leading into more homophily in India.

Birds of a feather flock together. Friend of a Friend ultimately becomes a friend because you spend more time together and tend to share more common interests together. Enemy of an Enemy tends to become a friend. This has not only been proved using our paper but it is seen is reality in literally every friendship network that exists. Look around, homophily is everywhere.

V. REFERENCES

[1]. Homophily: the urban history of an algorithm news articles

[2]. Carrington, P.J., Scott, J., Wasserman, S. (eds.): Models and methods in social network analysis. Cambridge University Press (2003)

[3]. Wasserman, S., Galaskiewicz, J.: Advances in social network analysis: Research in the social and behavioral sciences. Sage Publications (1994)

[4]. www.academia.edu.com

[5]. Social Networks course on NPTEL

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Document Analyser Using Naïve Bayes Classifier

Uma. N[1], Shreya Korada[2], Sweta Leena[2]

[1]Sr Assistant Professor, Department of Computer Science & Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]UG student, Department of Computer Science & Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

For well multi classes and dataset prediction with easy and fast method, Naïve Bayes is the well-designed algorithm to use. The using of Naïve Bayes, for analysing the emotions used behind the content of the document, is easy and active method. It is not only easy process but also cost-effective so all leman users also can use this application easily. The advantage is, it is user-friendly and takes very less time to analyse the document. It checks the sentiment of each word in terms of positive, negative and neutral words, which will help analyser to analyse the document, so instead of checking each word manually in the document, analyser can accomplish the work within very limited amount of time. The application has designed in such a way that it will be able to count the number of positive, negative and neutral words then according to the number of words it will provide the graph which will make the work of analyser easy to analyse, also automatically it will save all the negative, positive words in separate text file in another folder. So that if some document has more negative words and the graph is red in colour, so analyser will have the words in separate folder as prove.

About the application: - Entire application developed by software developer using python and machine learning, using spider editor. The interface of the application is GUI (graphical user interface) which will help the user to get the result in the form of frame, so that it will increase the interaction between user and application.

Keywords: sentimental, analyser, analysis, Naïve Bayes, document, python, machine learning, GUI, analyser, graphs.

I. INTRODUCTION

In day today life many people upload so many research papers, documents. Analysing each and every document manually, which is appropriate or not, which has more positive words, more negative words and neutral words is difficult and time consuming. So, this research will help to resolve the above problem. The main goal of this research is to analyse maximum number of documents in very less time, when it comes to analyse, this application will analyse which document has, how much percentage of positive, negative and neutral words.

Naive Bayes is combined of two words Naive and bayes where both the words containits own algorithm

and when both the algorithms added together it makes machine learning easier and faster.

Naive- According to the studies this has confirmed that the occurrence of a certain feature is independent of occurrence of other features. For example, if engine, wheels, handle used for identifying the vehicle, then two wheels, cylindrical engine and handle will be recognized as a bike. So, this example explains how each feature individually contributed to identify the bike without depending on each other.

Naive theorem: Naive Bayes methods consist of some set of supervised learning algorithm based on Bayes theorem. Naive is highly used for solving classification problems, like text classification problem.

It helps to make the machine learning language fast which will be able to do predication in a small amount of time. This makes the predictions based on probability of objects.

Bayes: Bayes theorem helps to update the probabilities of a predicted event. If the conditional probability is given then it is easy to find reverse probability also. So, Bayes theorem depends on conditional probability. There is a formula which used to find the probability

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

For example: - if we have to calculate the probability of taking a blue bike from the second show room out of three different showrooms' collection of bikes, where each showroom contains three different colour bikes like red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

II. LITREATURE

Jianqiang, Zhao, GuiXiaolin, and Zhang Xuejun [8] developed a word embeddings method based on large twitter data with the help of unsupervised learning by combining co-occurrence statistical characteristic and latent contextual semantic relationships between words in social media. The sentiment features of social media s were formed by combining the word sentiment polarity score and n-gram features in word embeddings. The sentiment classification labels were predicted by feature set which was integrated into Deep Convolution Neural Network (DCNN). The efficiency of word embedding method was validated by conducting experiments on five datasets when compared with existing techniques. The pre-trained word vectors used in DCNN had good performance in the task of TSA. While clustering the sentimental contents in large dataset, the computational time becomes a bit high.

M.Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, [12] proposed a hybrid classification framework to overcome the issues of incorrect classification. The input text was passed through the first two classifiers such as emoticon and slag, after applying the preprocessing stage. In the final stage, SWN based and domain specific classifiers were applied to classify the text accurately. A limitation of the approach was the lack of automatic scoring of domain specific words without performing a lookup operation in SWN, which may increase the classification accuracy.

F. Bravo-Marquez, et al., [2] implemented a method for opinion lexicon expansion for automatically annotated social media s from three types of information sources such as social media s of emoticon-annotated, hand-annotated and unlabeled social media s. The domain-specific problem was tackled by transferring the method into annotation approach for unlabeled social media s.

III. METHEDOLOGY

Dataset is a collection of data. It can be in tabular format or in text format. If it is a tabular format then

every column represents a particular variable and each row correspondence to a given record of data set. There is various type of datasets

- -Numeric data sets
- -Bivariate data sets
- -Multivariate data sets
- -categorical data sets Working of dataset: -

According to the definition the dataset is collection of numbers or information, which are related to the topic. The three ways to work with dataset includes mean, median and mode.

Mean: - We can find the mean of a dataset by finding the average of the dataset.

Median:- The middle value of a dataset can be considered as the median of a dataset.

Mode:- Mode is the value, which occurs repeatedly in dataset. Or in other words, mode is the number or value that occurs most often in the dataset.

Main Project working principle

The dataset which is used for this research is co relational- data set.

Co relational data set:-

The set of data which are related to each other indicates co relational data sets, generally collection data sets work on prediction of correlation between the things, and generally co relational data sets classified into three types.

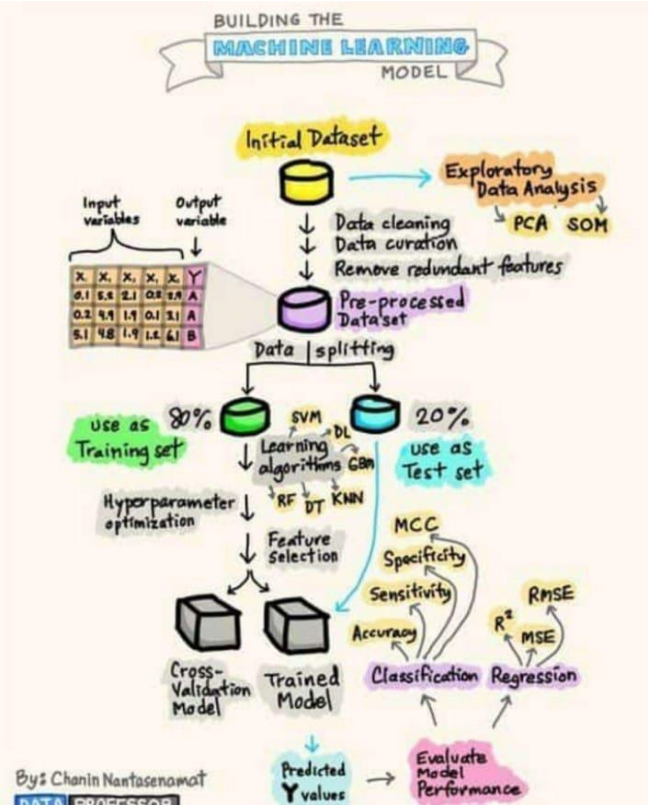


Fig. 1: working of dataset information

- Positive correlation
- Negative correlation
- Non-zero correlation

The main work of document analyzer is to analyze the document based on the positive, negative and neutral words. So machine learning makes the thing easy and helps to analyze the document faster and it has the ability of learning things by itself so if any new words come in future, then it will be able to learn it and can recognize whether the data/word is positive, negative or neutral. This research paper has designed in such a way it reads the data and matches with the database it will analyze each and all words present in the document so the runtime of the project depends on the words present in the document. The best ability of this project is it can store the words in separate folder and also it can display the words which are separated with comma and it is able to count the number of positive,

negative and neutral words. The graph also will be shown to the user based on the analysis.

```
positive_vocab = ['awesome', 'liked', 'outstanding', 'fantastic', 'terrific', 'good', 'nice', 'great', ':)']
neutral_vocab = ['book', 'the', 'sound', 'i', 'it', 'was', 'is', 'actors', 'did', 'movie', 'know', 'words', 'parts',
negative_vocab = ['bad', 'terrible', 'useless', 'hate', 'worst']
```

Table 1: Graphs according to the word range

Words	Word quality level	Graph color
Adaptable. Adventurous. Amazing. Amiable.	Positive	GREEN
No. Not. None. No one.	Negative	RED
disinterested, even-handed, fair, impartial,	Neutral	BLUE

IV. IMPLEMENTATION RESULTS AND DISCUSSION

This projected is implemented using NaiveBayesClassifier by importing NLTK library. GUI screen with buttons and text messages is developed by loading TK modules. Good part of this project, the concept of segregation of inputs can be separated (positive & negative in this case) and this concept can be used in many applications. As we discussed in the earlier section, the GUI screen is developed using TK modules. The button widgets with text messages, images etc. are embedded in main screen which allows user to click and go to the next steps.

From the above code, it can be seen that TK modules are imported from TKinter libraries. Also the button widget with “Welcome to NEW HORIZON” message is embedded in the GUI main page. New horizon image page is also placed by using PhotoImage() in-built function.

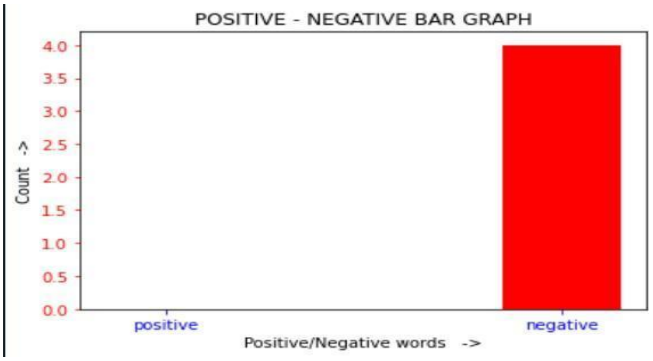


Fig.2: GUI Screen Code

Mainloop() function helps to run the UI code, which can be seen in above code snippet

Table 2: Analyzer Classification Model

SL.NO	MODEL NAME	WORK OF MODEL
1	RULE-BASED MODEL	Applies hand-crafted rules to establish the platform tag
2	AUTOMATED SYSTEM MODEL	System uses machine learning algorithm from past observaion
3	HYBRID SYSTEM MODEL	Hybrid system is combined of both rule-based and machine learning based
4	DECISION TREE MODEL	tree algorithm is a data mining induction technique that recursively partitions a data set of records using depth-first greedy approach

Table 2: .

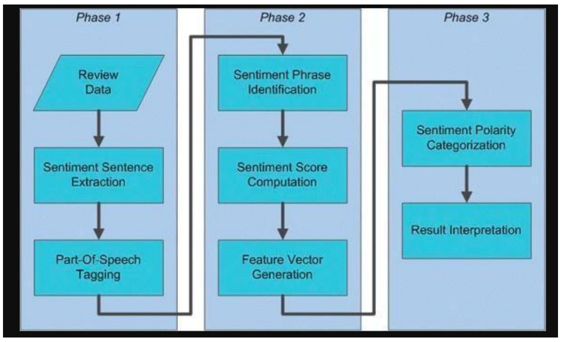


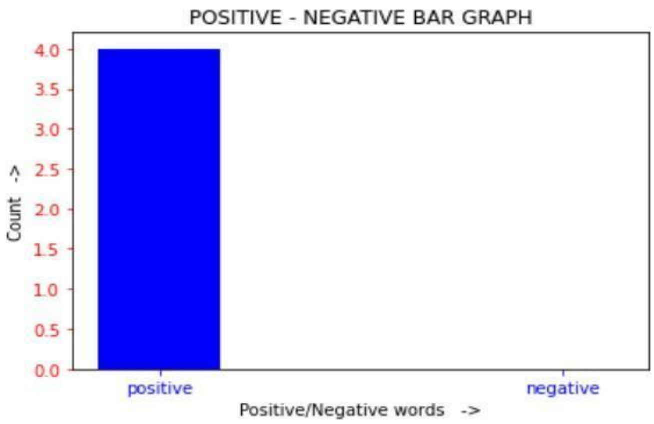
Fig. 3: working of document analyzer

V. OUTCOME AND DISCUSION

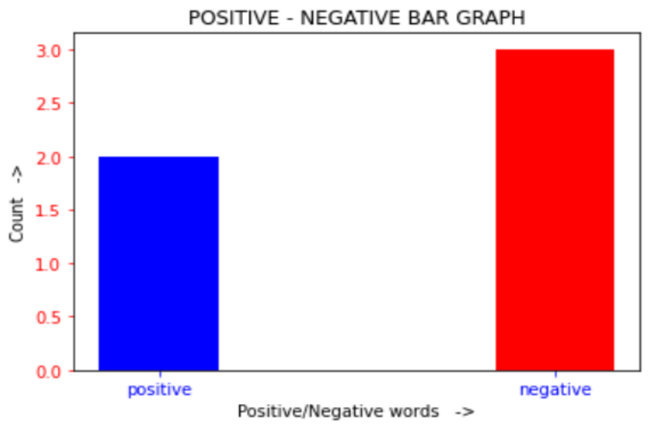
The 1st part of the project is GUI (Graphical user interface) based, which has a welcome page or a home page and a button "click to proceed", after clicking the button application will start analysis for the given document. After analysis the application will show the words, which are present in the document then it will give the count of the positive, negative, neutral words. If the positive words are more than the negativeand neutral words, according to rule- based model it will give the document positive tag and will give the graph based on that, the color of the graph will be green in color.

Graph3.0:- negative

Case:-1 The document which has upload through the application, has more negative words so the graph color is red.



Graph3.1:- Positive



Graph3.2:- mix of positive and negative

Case:-3 After seeing the positive and negative percentage analyzer can figure out how much is the percentage of neutral words, positive words and negative words

Case:-2 The document which has uploaded through the application, has more positive words so the graph color is blue...

VI. CONCLUSION

Document analysis or paper analysis is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online document reviews. A sentiment analysis working process has been proposed how exact the analysis process works. The process and the steps.

As we have seen, the "Document Analyzer" can analyze documents (text files). However, we can also analyze html files, excel spread sheets, CSV files etc. We can also analyze the sports columns, business news etc. based on our business requirements.

VII. REFERENCES

[1]. Alsmadi, and Gan KengHoon. "Term weighting scheme for short-text classification: Twitter

corpuses."Neural Computing and Applications (2018): 1-13.

[2]. F. Bravo-Marquez, Eibe Frank, and Bernhard Pfahringer. "Building a Twitter opinion lexicon from automatically-annotated social media s."KnowledgeBased Systems 108 (2016): 65-78.

[3]. Himja Khurana, and Sanjib Kumar Sahu. "Bat inspired sentiment analysis of Twitter data."Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore, pp. 639-650, 2018.

[4]. H. Shirdastian, M. Laroche, and M.O. Richard, "Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter," International Journal of Information Management, 2017.

[5]. Himja Khurana, and Sanjib Kumar Sahu. "Bat inspired sentiment analysis of Twitter data."Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore, pp. 639-650, 2018.

[6]. H. Ameur, Salma Jamoussi, and Abdelmajid Ben Hamadou. "A New Method for Sentiment Analysis Using Contextual Auto-Encoders."Journal of Computer Science and Technology 33.6 (2018): 1307-1319

[7]. H. S. Manaman, S. Jamali, and A. AleAhmad. "Online reputation measurement of companies based on user-generated content in online social networks."Computers in Human Behavior 54 (2016): 94-100.

[8]. Jianqiang, Zhao, GuiXiaolin, and Zhang Xuejun. "Deep convolution neural networks for Twitter sentiment analysis."IEEE Access vol. 6, pp. 23253- 23260, 2018.

[9]. Luis Terán, and José Mancera, "Dynamic profiles using sentiment analysis and twitter data for voting advice applications."Government Information Quarterly (2019).

[10]. M. Birjali, A. Beni-Hssane, and M. Erritali, "Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks," Procedia Computer Science, vol.113, pp.65-72, 2017.

[11]. M. Daniel, R.F. Neves, and N. Horta, "Company event popularity for financial markets using Twitter and sentiment analysis," Expert Systems with Applications, vol.71, pp.111-124, 2017.

[12]. M.Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme", Expert Systems, vol. 35, no. 1, e12233, 2018.

[13]. R.C. LaBrie, G.H. Steinke, X. Li, and J.A. Cazier, "Big data analytics sentiment: US-China reaction to data collection by business and government," Technological Forecasting and Social Change, 2017.

[14]. T. Singh, and M. Kumari, "Role of text preprocessing in twitter sentiment analysis," Procedia Computer Science, vol.89, pp.549-554, 2016

[15]. V. Vyas, and V. Uma, "An Extensive study of Sentiment Analysis tools and Binary Classification of social media s using Rapid Miner," Procedia Computer Science, vol.125, pp.329 -335, 2018. W. Xu, L. Jiang, An attribute value frequency -based instance weighting filter for naive Bayes [J]. Journal of Experimental & Theoretical Artificial Intelligence 31(4), 225 –236 (2019)

[16]. Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," Knowledge -Based Systems, vol.145, pp.207 -218, 2018.

[17]. http://www.jcreview.com/admin/Uploads/Files/61adf e235a6725.71379760.pdf

[18]. https://towardsdatascience.com/how -to -build - a - machine -learning -model -439ab8fb3fb1

[19]. https://www.techtarget.com/searchenterpriseai/feature e/How -to -build - a -machine -learning – model.
[20]. https://study.com/academy/lesson/data -set -in - math - definition -examples -quiz.html
[21]https://parade.com/1241177/marynliles/psitive-words/

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Decentralizing The World Using Blockchain

Naresh H Suthar[1], Joel Jacob Stephen[1], Keshav Krishna Kumar[1], J Mukundh[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper fills in the gaps in our knowledge about Blockchain and Web 3.0. The new technologies in this ever-evolving field and the underlying risks and opportunities offered by the advent of blockchain technologies along with the rise of Web 3.0 are some of the topics discussed. Those interested in these technologies and researchers and academics interested in fields such as sociology and cyberculture will find true meaning in this paper. It's no longer about bitcoin or cryptocurrency, there is so much more that lies beneath, for we know but the tip of the iceberg.

Keywords: Blockchain, Web 3.0, Metaverse, Decentralized.

I. INTRODUCTION

The problem that a blockchain solves is "consensus." It transforms the concept of trust by integrating components that can cause financial upheaval. As a result, currencies are the first concept that may be implemented in such technology, although this is merely the starting point. At first, interest in the bitcoin digital currency was strictly academic, and bitcoin exchanges were limited to a small group of people who were more interested in the cryptography qualities of bitcoins than the actual bitcoin value. Bitcoins are now used to buy and sell actual products and services in the same way that traditional currencies are.

The distributed infrastructure where all transactions are recorded is the Bitcoin system's most distinguishing feature. A user needs an alphanumeric code known as an address in order to transmit and receive bitcoins. An address can be viewed as a bank account number and used to send money. A public key yields public information in the form of an address. Because no personal information is stored on a blockchain, the Bitcoin system provides pseudo-anonymity. The consensus mechanism enables agents to transfer "value" without the involvement of a third party, ensuring that the source owns the value being transferred and that the recipient receives (or does not receive) the value being transferred. This has been eliminated. Anyone can see the entire transaction history, and it cannot be modified.

The introduction of blockchain technology ushers in a new age of the web, which we refer to as Web 3.0. The early days of the Internet were dominated by information delivered via static websites with no opportunity for interaction. It was primarily created by information portals with flat material that allowed visitors to "just" read but not submit any comments, reviews, or feedback. The capabilities to communicate, share information, add content, and trade data define

Web 2.0, or the second stage of the World Wide Web's evolution. This age, also known as participatory, allows all users to participate, create online material (user-generated content), and interact with other users effortlessly (usability). Web 3.0 is built on decentralization, with no one point of control or profit center. The blockchain allows value to be transferred without the need for a profit center or monopolistic service providers. While social media allowed users to share information, it concentrated control in the hands of a few private actors (creating digital oligarchies with social media companies, peer-to-peer ridesharing, and peer-to-peer hospitality networks), blockchain technologies allow for the creation of decentralized networks with no centralized points of control. One of the technology's most disruptive features is its ability to operate on a decentralized basis without a central profit center in charge of managing (and exploiting) the network. Blockchain technology enables the secure transmission of data, assets, and services. Without the use of a third-party intermediary, such as banks or other financial organizations, blockchain technology enables the secure transfer of information, assets, and money.

Metaverse in the blockchain is transforming the world of social media platforms. Metaverse provides endless interconnected virtual communities using virtual reality headsets, augmented reality glasses, and smart apps, whether you are going to be concerned, trying on clothes, working in an office, or going to work.

II. METHODS AND MATERIAL

The process used to write this paper involves essentially combining various other resources on the blockchain, Web 3.0, and the Metaverse to be able to establish some basic grounds on what each of these terms actually means, and how well are people aware of these terms. The latter was done using a survey created by us, in order to truly comprehend on what level people of various demographics aware of these are rapidly changing technologies.

III. RESULTS AND DISCUSSION

After thorough research and studying, we have put together information on each of the aforementioned topics in order to get a clearer understanding of each.

Blockchain

Blockchain technology, decentralized apps, and distributed ledgers are becoming the foundation of our lives. Blockchain is no longer just about bitcoin or cryptocurrencies in general. Instead, it is now seen as a disruptive, revolutionary technology that will have a significant impact on multiple aspects of our society. The revolutionary power of this technology compares with the revolution sparked by the dawn of the World Wide Web and the Internet in general.

In this paper, we explain these technologies and how they will affect the social, technological, and economic landscape going forward. We also discuss the Metaverse which is a post-reality universe that aspires to be a perpetual and persistent multiuser environment that merges physical reality with digital virtuality.

The problem solved by a blockchain is “consensus”. It revolutionizes the concept of trust, introducing elements for generating disruption in the financial sector. Currencies are therefore the first concept that can be implemented upon such technology, but this is only the premise. Satoshi Nakamoto conceived the Bitcoin electronic cash system in 2008 with the aim of producing digital coins whose control is distributed across the Internet rather than owned by a central issuing authority such as a government or a bank. It

became fully operational in January 2009, when the first mining operation was completed, and since then it has continuously seen an increase in the number of users and miners. In the beginning, the interest in the bitcoin digital currency was purely academic, and the exchanges in bitcoins were limited to a restricted elite of people more interested in the cryptography properties than in the real bitcoin value. Nowadays bitcoins are exchanged to buy and sell real goods and services, as happens with traditional currencies.

The main distinctive feature introduced by the Bitcoin system is the distributed infrastructure where all the transfers are recorded. To send and receive bitcoins, a user needs an alphanumeric code called an address. An address can be seen as a bank account number and can be the recipient of funds. An address is a public information derived from a public key. No personal information is recorded in a blockchain, and for this reason, the Bitcoin protocol offers pseudo-anonymity. The consensus mechanism allows agents to transfer "value" without having a third party involved in the process, which guarantees that the source actually owns that value that it wants to transfer and which guarantees that the recipient receives (or not) the value being transferred. The elimination of this third trusted party is a major breakthrough. If we think about how banks work at the moment and what they actually do, it is immediately clear that banks match the definition of trusted third parties. The bitcoin blockchain allows the transfer of value without a third party. The disruptive potential of the bitcoin consensus algorithm is enormous. The fact that all the transactions are public and it is not possible to delete them is the key that lets the consensus algorithm work. The whole transaction history (from the first that occurred) is accessible by anyone (any agent who wants to check what happened from the genesis), and it cannot be changed.

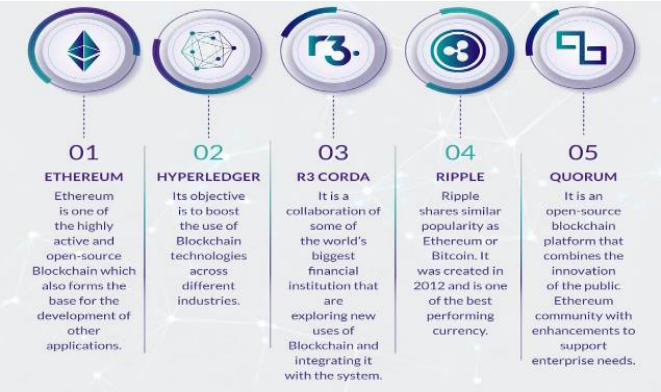


Fig. 1: A representation of the various Blockchain Platforms

Web 3.0.

The advent of blockchain technology brings a new era to the web, what here we define as Web 3.0. The first era of the Internet was mainly characterized by information carried by static websites without any possibility of interaction. It was primarily made by information portals with flat data where users could "only" read and were not allowed to add any comments, reviews, or feedback.

Web 2.0, or the second stage of the World Wide Web's evolution, is characterized by the possibilities to interact, share information, add content, and exchange data. This era, also known as participative, gives the possibility to all users to participate, generate content online (user-generated content) and easily interact with other users (usability).

Web 3.0 is based on decentralization, without points of control and unique profit centers. The blockchain enables the transfer of value without a center of profit or monopolistic service providers. While the advent of social media allowed the exchange of information among users but kept the control among a few private actors (generating digital oligarchy with social media companies, peer-to-peer ridesharing, peer-to-peer hospitality networks), blockchain technologies allow the possibility of creating decentralized networks without centralized points of control. From this stems

one of the disruptive aspects of this technology that will enable the operation of a decentralized system without any central center of profit in charge of coordinating (and taking advantage of) the network. Blockchain technology allows the secure transfer of information, assets, and money without a third-party intermediary, such as banks or other financial institutions (Swan, 2015: 15). These third-party intermediaries are not limited to banks, but it also includes the economic platforms of the sharing economy and Web 2.0, which make a profit from each transaction, and popular social media platforms, which make profits using users' data. A blockchain can be used also as a backbone infrastructure for running smart contracts, particularly decentralized applications which can be seen as computer programs executed by participants in a blockchain. Smart contracts are an additional disruptive factor and have gained tremendous popularity in the past few years, to the point that billions

of US dollars are currently exchanged every day through such technology. However, since the release of the Frontier network of Ethereum in 2015, there have been many cases in which the execution of smart contracts managing Ether coins lead to problems or conflicts. Smart contracts rely on a non-standard software life cycle, according to which, for instance, delivered applications can only be updated or bugs resolved by releasing a new version of the software. Furthermore, their code must satisfy constraints typical of the domain, like the following: they must be light; the deployment on the blockchain must take into account the cost in terms of some crypto value; their operational cost must be limited, and they are immutable since the bytecode is inserted into a blockchain block once and forever.

Some of the characteristics of Web 3.0 are as follows:-

Semantic Web:

The Semantic Web enables machines to more effectively understand meaning and emotions, improving user experience and easing connectivity. To put it another way, it makes use of Artificial Intelligence (AI) to improve data processing.

3D Graphics and Spatial Web:

Another feature that distinguishes Web 3.0 from Web 2.0 is 3D graphics. The three-dimensional design allows for the creation of a more realistic and engaging cyberworld, which expands business opportunities. Some industries, including architecture, online games, real estate, e-commerce, tourism, and others, have already used 3D graphics to promote their products and services. Spatial Web, on the other hand, is thought to be the future of Web 3.0, blurring the line between the real and digital worlds. Some of its technologies, such as Virtual Reality, are already being used in organizations and businesses to assist users in visualizing objects and processes.

Metaverse.

The following are some of the Features, Limitations, and Current players in the industry in The Metaverse.

Features

The metaverse will be an ecosystem of shared virtual worlds:-While virtual worlds do exist today, the majority of them are unconnected, and users cannot freely navigate between them in the same way that they can from one website to another. In most cases, assets purchased in one game typically cannot be used or sold in a different game. Items like digital apparel and accessories will be able to be shared across worlds in the metaverse. The rise of technologies such as non-fungible tokens (NFTs) may help users use products across different worlds.

The metaverse will fuse the real world with the digital world:-Creators will be able to replicate real-world spaces or things in a virtual medium, allowing viewers to feel as if they are physically present in the space rather than simply seeing it through a computer browser. AR technology can also be used to insert virtual things into the actual environment. Companies and creators will find new ways to blend the physical and digital to create experiences that feel more real and connected with the typical consumer's daily life as AR and VR technology develops.

The metaverse will provide new ways for people to connect and interact with each other in a virtual space. Many in the business believe that the metaverse has the potential to change how and where people interact. The pandemic has demonstrated how critical it is to establish ways for people to meet and interact with one another through a virtual medium. People will be able to meet, chat, and participate in activities and events together using AR and VR, even if they are physically separated. Even if travel becomes more accessible, families separated by countries or continents will see the metaverse as a way to stay in touch in a way that feels as if they were beside them. Virtual spaces can democratize access to events, education, and even job prospects by eliminating the physical obstacles and constraints of the actual world.

Limitations

Companies are attempting to overcome present network restrictions, improve device capabilities, and build common standards and protocols to bring the metaverse to life on a broad scale. Some of the limitations to the advancement of metaverse are:

Network Limitations:

The metaverse will require persistent connections that update in real-time with high precision and a quick data delivery time in order to function. The metaverse must also enable a diverse set of applications and experiences across all platforms while catering to a vast number of users. This is a massive computational difficulty that goes against the Internet's basic design or aim. The three main criteria used to assess network quality are bandwidth, latency, and reliability.

All three of these network restrictions must be overcome in order to bring an immersive metaverse to the masses.

Hardware:

Hardware restrictions, according to most industry insiders, will be the biggest roadblock to creating the metaverse. Users will need to use AR/VR devices to access the metaverse and have a completely immersive experience. Current headsets, on the other hand, are costly, cumbersome, and lack the graphics quality and computing capacity needed to produce realistic virtual worlds or objects. For consumers interested in participating with the metaverse, this provides a significant barrier to entry. Continued advances in the underlying hardware components, such as processor chips, graphics chips, and light engines, will be required to reduce the cost and size of AR/VR systems.

Adoption of interoperable standards:

Interoperability is one of the reasons why people believe that blockchain technology will power the future metaverse and that decentralization would make it easier for developers of different platforms to connect and collaborate with each other to provide a smooth experience to users.

The internet of today is governed by a set of common standards and protocols that enable anyone to establish websites, transfer files, and interact with it in other ways. In the metaverse, standards and

protocols will need to be far broader and more robust. Furthermore, due to the importance of interoperability and real-time synchronous experiences, a common set of standards will be required.

IV. METAVERSE ECOSYSTEM PLAYERS

The Metaverse Ecosystem currently comprises many players as shown below:-

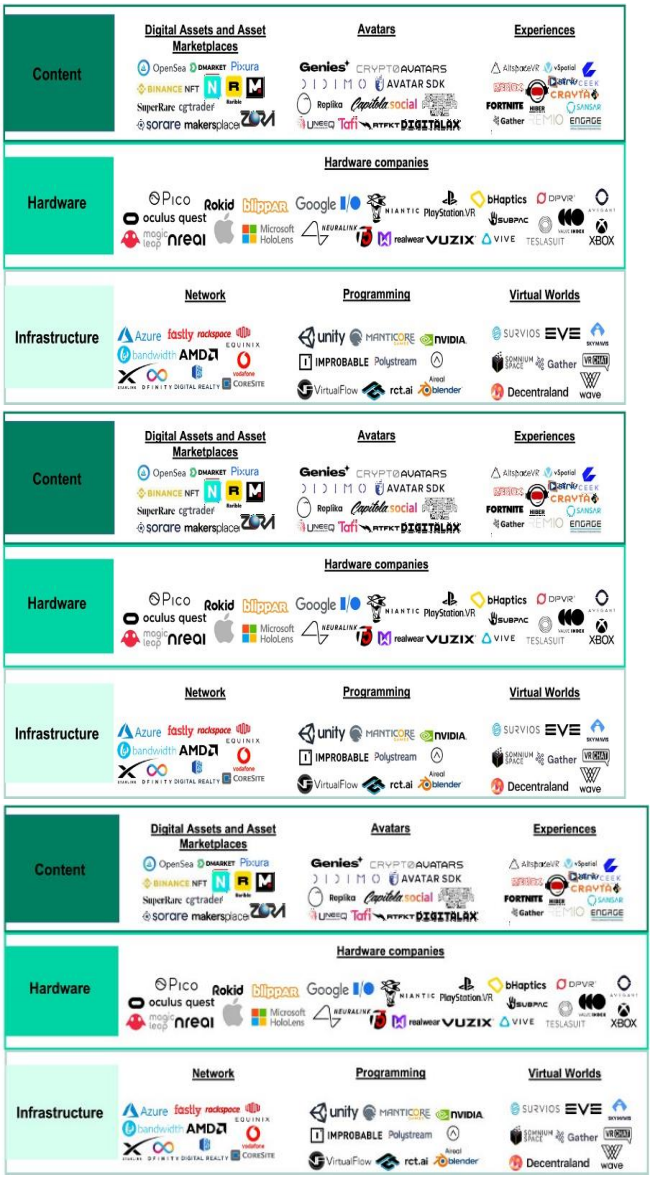


Fig. 2: A representation of the various Metaverse Players currently in the market, primarily in the fields of Content, Hardware, and Infrastructure.

Survey Results

The Survey performed included asking people of different ages a few questions on Blockchain and their awareness of the same.

The following are some of the results obtained according to the differentiating demographics:-

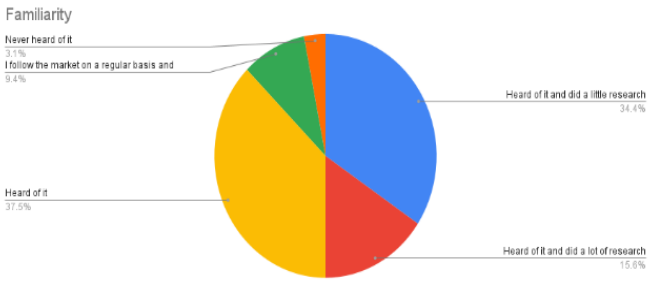


Fig. 3: A representation of the familiarity of Blockchain

While most of them had heard about it and did moderate research on the same, there were still a few who were not aware of these technologies.

RESULTS-1

37.5% - Heard of Blockchain.

34.4% - Heard and did a little research.

15.6% - Heard and did a lot of research.

9.4% - Follow the market on a regular basis.

3.1% - Never Heard of it.

We then asked them if they thought Blockchain and its technologies would help us in the future, the results we got for the same are as follows:-

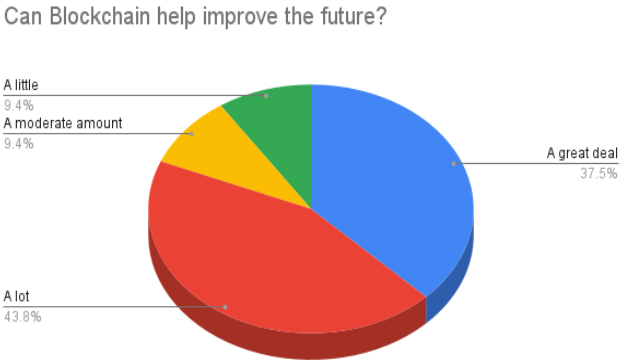


Fig. 4: A representation of the Blockchain's role in the future

RESULTS-2

43.8% - Help a lot.
37.5% - Help a great deal.
9.4% - A moderate amount.
9.4% - A little.

V. CONCLUSION

The paper helped us in understanding the roles of these ever-evolving technologies and how people from various demographics are aware of the same.
We were able to establish certain points through our research and able to give what is just a gist of the actual vast world of Blockchain and its technologies.
The Survey performed also helped us understand what exactly people think of this world of Blockchain and how there is so much more to be explored in this field.

VI.REFERENCES

[1]. https://ieeexplore.ieee.org/abstract/document/8441990
[2]. https://arxiv.org/abs/1910.06092
[3]. https://www.mdpi.com/2624-6511/5/1/
[4]. Wikipidea/Blockchain

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

WEBGEN 3.0

T J Tejas Maladkar[1], Milina M[1], Serena M[1], Rahmath Mohis[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

In this era of technological evolution, Web 3.0 has had a significant impact. It includes the semantic web, decentralization of block chains, artificial intelligence (AI), and the Internet of Things (IoT). Blockchain isn't just about bitcoin or cryptocurrencies.It is a revolutionary technique that will impact many aspects of our lives. Blockchain's revolutionary power compares with the revolution of the World Wide Web and the Internet. It can be a game-changer for introducing the next level of sharing information; sharing value. This study presents the relationship between current technologies (Web 1.0, 2.0, and the third generation of the web, which is Web 3.0). Web 3.0's most recent evolution necessitates Web 2.0 technologies, semantic web, and artificial intelligence.

Keywords—Web3.0, Web2.0, Web1.0, Block chain, Semantic web, artificial intelligence, IoT, RDF
Introduction (Heading 1)

I. INTRODUCTION

The Internet has played a vital role in the growth and development of every aspect of our lives. The Internet has global broadcasting capability and is the most common medium of communication between people and computers.A British scientist named Tim Berners-Lee invented the World Wide Web (WWW) in 1989. In the early days, the Web was conceived and developed to meet the demand for automated information sharing between scientists in certain universities and institutes around the globe.

Web1.0 was the first stage of the evolution of the World Wide Web (WWW). Only content creators were able to access the Web 1.0. This website was mainly used for content delivery networks (CDN), which enables the showcase of particular pieces of information on the website. It contains certain directories that enable users to retrieve particular pieces of information

Later, the second stage of evolution took place, which is called Web 2.0. It is also called the "participative social web" and highlights user-generated content, usability, and interoperability for the end users. Interaction between the application and the end user was made possible using Web Users were also able to provide content instead of simply viewing it. People were allowed to publish articles and comments, and it became possible to create user accounts on different sites, which led to increased participation. There was a rise in web pages, surfing, and accessing web applications.

Web 2.0 is used in social media, like Facebook, Instagram, Whatsapp, etc., and content-providing

sites like Wikipedia, blogs, etc. Web 2.0 gave rise to what are considered the three layers of innovation: mobile, social, and cloud. Web 3.0, originally called the Semantic Web includes Blockchain, Artificial Intelligence, and the Internet of Things (IoT).

Polkadot founder and Ethereum co-founder Gavin Wood, in the year 2014, first coined the term "Web 3.0," which is also called the third generation of the World Wide Web (WWW). Users are now allowed to read, write and also execute the web, offering an integrated web experience. Machines were enabled to understand and catalog the data in a human manner. This new generation of the Web facilitates a world-wide data warehouse in which different formats of information are shared.

It is understood that the advancement of Web 3.0 will revolutionize the internet. Different organizations and some academic institutes use the technology of Web 3.0 for their business and academic and research purposes. Decentralized finance, often known as DeFi, is a component of Web 3.0 that entails executing real-world financial transactions on the blockchain without any help from banks or the government.

Web 3.0 mostly works on two things, and they are:

Decentralization: there is no involvement of a central authority to provide permission to post anything on the web. There is no kill switch and no single point of failure as there is no central controlling node. All this together leads to freedom from indiscriminate censorship and surveillance.Bottom-up design: the code was developed with total participation, without being written by a small group of experts.

Fig. 1: Web 3.0 Unique

II. RELATED WORKS

The idea of Web 3.0 was first suggested in 2006 by a reporter of The New York Times, named John Markoff. This segment aims to emphasize a few analyses that have been made to make Web 3.0 a reality in the future.

Need for Web 3.0

The immediate need for Web 3.0 was addressed by Anton Dziatkovskii, CEO and co-founder of Platinum Software Development Company, in his paper on May 15th, 2021. Today, mass censorship and de-platforming have been normalized, and cryptocurrencies have facilitated a brighter, less surveilled future, leading to Web 3.0. The idea of Bitcoins (BTC), a sovereign, leaderless, deflationary currency, has been able to counter centralized monetary power by being decentralized and secured by blockchain [1]. Web 3.0 (previously known as the Semantic Web) embraces the concept of an open internet in which data becomes portable and owned by the user.Thesemantic metadata would make it easier for users to connect to the content and services they require.To level the Web 3.0 experience, such content would be seamlessly distributed among devices and apps. An idea of what this Web 3.0 would look like and why certain significant hurdles of the

Web are worth overcoming is also discussed in Mr. Anton's work.
DOI: 353143518

Governance, Risks and Safeguards of Web 3.0:

The Web 3.0, once implemented, would need to be governed and safeguarded from certain risks. A proposal to achieve this was given by Mr. Rikus (Hendrik Jacobus) Bruwer and Mr. Riaan Rudman from Stellenbosch University, South Africa in May 2015. Technology is an important asset to many organizations to create revenue while controlling costs. However, the Web, as the world's fastest-growing publication medium, with a massive amount of unstructured data, presents numerous new options for businesses. Web 3.0 is thought to promote the creation of a global data warehouse in which any format of data may be shared and understood by any device over any network, posing both opportunities and hazards.Mr. Rikus and Mr. Riaan discuss the dangers of Web 3.0 and suggest possible protections to reduce these dangers to a manageable level [2]. Unauthorized data access and manipulation, autonomous action initiation, and the development of scripts and languages are all examples of risks.Control processes like encryption, access control, filtering, language and ontology development might help to mitigate these risks.

Web 3.0 Architecture

An idea of the architecture for Web 3.0 and the emerging time order was given by Hengjin Cai and Tianqi Cai, School of Computer Science, Wuhan University, China. The architecture for Web 3.0 is based on hashed interactions among user nodes that can transform bilateral trusts into collective time order [3]. This is the major achievement of blockchain technology, without the expensive Proof of Work or the questionable Proof of Stake. The high transactions per second, low storage costs, and flexibility in data management will provide a basis for a new paradigm in individual, corporate, and social governance, privacy protection, and personalization. Based on their work, the hacked interactions leading to intrinsic time order bring out the state with established rules and the state of autonomous interaction.

Web 3.0 and Linked Blockchain Data

A more practical version of Web 3.0 and linked blockchain data ["Sematic Blockchain"-Starting Point] was given by Héctor E. Ugarte and Rheinische Friedrich from Wilhelms Universität Bonn, Germany on June 1, 2017. Semantic web done right is Linked Data. The Semantic Web is an incomplete dream so far, but it is a futuristic Internet that may be composed and powered by blockchain networks, interconnecting data and meaning, thus allowing reasoning. Blockchain can be thought of as the longest path from the root of the tree to the leaf, serving as a uniform transaction history on which all nodes eventually agree.The concept of Resource Description Framework (RDF), being a family of W3C specifications, is defined as "it is a foundation for processing metadata" [4]. Some other terms mentioned in their work are ontology and Web Ontology Language. An ontology is a collection of explicit formal definitions of words (categories or concepts) in the domain and their relationships (structures or roles) between them. OWL (Web Ontology Language) is a language designed to represent complex and rich information about objects, objects, and relationships [4]. Héctor E. Ugarte R and Rheinische Friedrich also discuss the concepts of blockchain linking, smart contacts (a term coined by Nick Szabo and defined as a tool for automating human interactions), Ricardian Contracts, Trust Contracts, Domain Capture, Decentralized Storage,

and RDF), Semantifying the Blockchain, Semantic Blockchain Network, and Supply Chain Scenario:Decentralized Supply Chain Application (DeSCA).
DOI: 315619465

Emergence of Web 3.0

The emergence of the exciting new Web 3.0 and the future of open educational resources was put forward by Jose Bidarra and Vitor Cardoso, UniversidadeAberta, Lisbon. The perception of video games is of two types: one is that video games can teach skills like analytical thinking, development of strategies, creativity, and so on. The other is that e-learning can benefit from video games. Following the mainstream collaborative and social WWW 2.0 (MySpace, Pandora, and YouTube), the pursuit of WWW 3.0 (Second Life, Joost, and VRML/X3D worlds) appears to be marked by this combination of human-like avatars, smart avatars, and rich multimedia features that live happily within interactive 3D environments [5] .These researchers explore new ways to make e-learning more effective in the future realm of open educational resources.
DOI: 228722095

Technologies and Social Media Services Towards Web 3.0

The idea of technologies and social media services moving towards Web3.0 was presented by Neil Y. Yen, AgustinusBorgyWaluyo, Chengcui Zhang, and James J. Park in a research setting. The time to change the web and its specifications has begun to promise to improve the way information is retrieved, managed, and delivered, so we can provide better support, through different types of services, operations, etc., in our daily lives. Successfully compiling a special article "Social Media Services and Technologies towards Web3.0" in Multimedia Tools and Applications, it is possible for reviewers (and interesting compendiums) who have been involved in this or related sites to agree that advanced knowledge, empowerment, and benefits are brought to this promising research space [6].

Evolution of the World Wide Web

The evolution of the World Wide Web from Web 1.0 to Web 4.0 was proposed by SarehAghaei, Mohammad Ali Nematbakhsh, and HadiKhosraviFarsani, University of Isfahan. The researchers provide us with an overview of Web 1.0, Web 2.0, Web 3.0, and Web 4.0. Web 1.0 was introduced as a tool for thought, Web 2.0 as a means of communication between people, and Web 3.0 as a digital network technology to support human interaction. San Murugesan described Web 2.0 as the second phase in the web's evolution in which technologies, services, development approaches, and tools of Web 2.0 were introduced in detail [7] . A comparison of the characteristics and technologies of the web generations and the progress of the web during the past two decades is also discussed in their work.
DOI : 10.5121/ijwest.2012.3101 1

The Ocean Protocol

Data is reckoned to be an exploited, illiquid asset. In today's world, the amount of data being produced daily is immense, but it's not utilized. Private data is not under the owner's control and it is used for free by companies, who then sell it without the users' knowledge. The user only realizes that their data has been sold when their identity is stolen or through the spam email folder. Ocean Protocol is the first to use the NFT concept to produce a data business where users can turn their data means into data NFTs [8] . The user can tokenize, control access, and monetize their data just by having a Web 3.0 wallet and an

internet connection. This remarkable initiative, Ocean Protocol, was awarded a technology pioneer in 2021 by the World Economic Forum.

A beta version of Ocean Protocol was launched in late 2020 when the marketplace experienced an unexpected demand for tokenized datasets. Very soon, The Ocean Protocol can be expected to launch "Ocean Data Market v4" and the tools for creating data NFTs may soon be unveiled. Ocean Farming is a data farming programme that will incentivize staking on premium-quality data assets with the protocol's native token, Ocean. Through NFTs, the data owners can leverage the Ocean protocol to generate revenue when creating NFTs of their datasets and also allow others to access the data by creating sub-licenses in Datatokens.

Examining the Metaverse

Online marketing has proven to be profitable for retailers in terms of revenue generation. Consumers spend most of their time and money on e-commerce sites and, now, m- commerce sites. The Idea for Moving Visual Marketing to Reality: Examining the Metaverse and Favorite Facts in Online Shopping Information provided by Esther Swilley reveals how consumer behavior has changed with the advent of the Web [9]. Research shows that Cyber Monday, the Monday after Thanksgiving, has exceeded the previously claimed Black Friday (Internet Retailer 2011). However, offline shopping is preferred over online shopping. Brick and mortar creates interaction, product knowledge, and excitement. Shopping for catalogs and televisions offers the convenience of home shopping. Although many consumers turn to their tablet computers to shop online, the shopping experience on tablets should compete with other online activities. In order to meet the needs of consumers, retailers need to build more robust online shopping information. This will improve consumer knowledge in both mobile and electronic trading. The Metaverse concept, when used, can allow users to experience visual purchases similar to real- world purchases.

DOI: 10.1007/978-3-319-24184-5_163

Starting Point of the Metaverse

The MetaHuman Creator's Proposal, The First Metaverse Site, was launched by Zhixin Fang, Li Bai Cai, and Gang Wang on February 11, 2021. Epic Games unveiled the first look at MetaHuman Creator, a new browser-based application that will give anyone the power to make a digital photorealistic bespoke person, fully bound and complete with hair and clothes, in just a few minutes[10]. The creator of Metahuman created digital avatars with a precise workflow, meaning that the Cyborg era is near.

DOI: 10.1109/ISCTIS51085.2021.00040

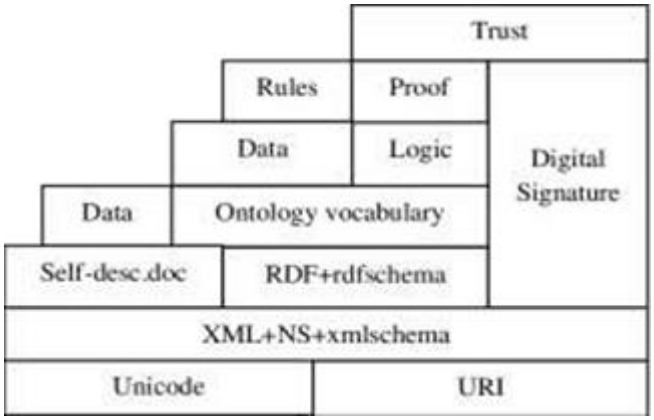
III. DIFFERENCES BETWEEN WEB 1.0, 2.0 AND 3.0

Web 1.0	Web 2.0	Web 3.0
Web 1.0 portrayed an environment where the information and the data was static. It displayed no interaction between the information and the consumer.	Web 2.0 is characterized by the possibilities to interact, share information, add content and exchange data.	Web 3.0 has the ability to formulate meaning from the context in which the information is published, and can make use of unstructured data more intelligently.
The protocols associated with this generation were Hypertext Transfer Protocol (HTTP) and Hypertext Markup Language (HTML).	Web 2.0 is about XML Web2.0 used most of the technologies present behind the internet such as XHTML standards, sheets, content syndication, AJAX and flash etc.	Web 3.0. OWL creates the opportunity for machines to adopt these characteristics and, through RDF linking, enables a Web where information is categorized by machines in a meaningful manner and in a format that can be used by other scripts.
Web 1.0 was a place where people could publish all their information for others to view.A paper named NanfangDaily still works on the Web 1.0 concept.	Web 2.0 emerged as the "web of people", it introduced users to collaborate with content creation and alteration Here users are asked to log in and then they are taken from one page to the other via these intralinks.	Web 3.0 formulates meaning from the published context to work more intelligently.
It is described as a web of documents. It has Sliced data.	Web 2.0 supports Light interlinked data	Web3.0 web oriented architecture and world wide database.
It has very Low-Portability. It could be only operated through the computing equipment	It has medium portability.it can be accessed through mobiles and computers.	It is highly portable. It can be connected to mobile, computers ad smart electronics
Professionally developed stand-alone applications	User-developed open applications	User-developed smart applications
Web 1.0 has Syntax aware basic browsing and search capabilities	Syntax-aware advanced browsing and search capabilities	Content(semantic)-aware and context-aware next-generation browsing and search capabilities
Point-to point/hub & spoke architecture	Service-oriented architecture(SOA)	Web oriented architecture(WOA) and internet of things

IV. DESIGN AND IMPLEMENTATION

Semantic Web Architecture

The basic semantic web architecture has Unicode and URI as its bottom layers. Each character is represented uniquely, and each resource type has a Uniform Resource Identifier (URI). The next layer is XML Extensible Markup Language. This includes other sets of standards as well, such as schemas and namespaces (NS), which are used to form common ways to shape the data on the web. XML is developed for the top layers of the semantic web. Next comes the Resource Description Framework (RDF). RDF is one type of data model that uses URIs for identifying web-dependent resources. This describes the relationships that exist between the named properties and their values. The RDF schema contains a predefined type of system for its models. Classes and the properties of the resources are described by the RDF schema. job



The ontology layer describes the properties and relations between them. In other words, it is a collection of terms that are used to describe a domain with the ability to infer. The layer above the ontology layer is the logic and proof layer. It creates new inferences through a system that uses automatic reasoning. Using these reasons, the agents can make a decision. The final layer of the architecture is trust. This is to provide some assurance of the quality of the information provided on the web. It also assumes a degree of confidence in the resource.

Semantic is an extension of the web, which contains data descriptors on existing data. It is created in a manner that computers can read the data directly. As a result of this, connectivity will be increased. If every data has a meaning in machine readable form, it gives better reasoning over the data and working on heterogeneous data is also possible. Data can be shared or reused over the common framework provided.

Algorithm: Possible Implementation of Semantic Web:
Input: User Entered Key words
Output: Mapped clustered/categorized result with user data Begin

1. User identification
2. Keyword Entered
3. If keyword == ontology
4. Categorize Hits, C
5. While C>0 do
6. If user, U ∈ UTHM database
7. Map RDB to RDF
8. Initialize n = 1
9. If Field data F == C
10. Rank = n
11. Else
12. Rank = n+1
13. End if
14. Else
15. Rank = n+2
16. End if
17. Done End

V. RESULTS AND DISCUSSION

It is found that Web 3.0 will give rise to a different era, with respect to technology. The next generation of the internet will have portable and connected data, due to the semanticity of the web. It will be capable of understanding and interpreting data like a human

would. Metaverse and Metahuman would no longer be fiction, as a result of virtualization techniques. The overall internet experience will be sensational. Online shopping and other activities would be enhanced due to the interactive visual purchases provided by Web 3.0. Online Education would gain more popularity and would be exceptionally enjoyable. Web 3.0 is that stage of the internet where there would be a line as thin as paper between reality and digital fiction.

VI. CONCLUSION

In this paper, an attempt has been made to understand the concepts involved in Web 3.0. Through this paper, one can improve their knowledge of the concepts of the World Wide Web, Artificial Intelligence, and Decentralization of blockchain technology, which is believed to be the solution to many existing problems. The upcoming wave of Web 3.0 goes far beyond our imagination, assuring us of more secure, user-driven services. Now that Web 3.0 is in the making, we are not far from the next generation of the web. Web 4.0 is gaining only slow recognition. The assumptions made for Web 4.0 include that people will interact with intelligent machines often and that a digital alter ego of every person is not that far away. It is also expected that it will be transparent and that privacy will be extinct.

VII. REFERENCES

[1]. Why Do We Urgently Need Web 3.0? by Anton Dziatkovskii. DOI: 353143518.

[2]. Web 3.0: Governance, Risks And Safeguards by Mr. Rikus (Hendrik Jacobus) Bruwer, Stellenbosch University, South Africa Mr. Riaan Rudman, Stellenbosch University, South Africa.

[3]. An Architecture for Web 3.0 and the Emergence of Spontaneous Time Order by Hengjin Cai and Tianqi Cai 1School of Computer Science, Wuhan University, China.

[4]. A more pragmatic Web 3.0: Linked Blockchain Data ["Sematic Blockchain" - Starting Point] by Héctor E. Ugarte R and Rheinische Friedrich from Wilhelms Universität Bonn – Germany. DOI: 315619465.

[5]. The emergence of the exciting new Web 3.0 and the future of Open Educational Resources by Jose Bidarra, Vitor Cardoso, UniversidadeAberta, Lisbon. DOI: 228722095.

[6]. Social Media Services and Technologies Towards Web 3.0 by Neil Y. Yen, Chengcui Zhang, AgustinusBorgyWaluyo and James J. Park. It is a Case study.

[7]. Evolution of the World Wide Web: From Web 1.0 to Web 4.0 by SarehAghaei (M.Sc. student), Mohammad Ali Nematbakhsh (assistant professor) and HadiKhosraviFarsani (PhD. Student) of software engineering in computer engineering department of University of Isfahan. DOI: 10.5121/ijwest.2012.3101 1

[8]. Ocean Protocol: A Web 3.0 Project Unlocking The Data Economy With NFTs by NavidLadaniFebruary 10, 2022.

[9]. Moving Virtual Retail into Reality: Examining Metaverse and Augmented Reality in the Online Shopping Experience by Esther Swilley. DOI: https://link.springer.com/chapter/10.1007/978-3-319-24184-5_163

[10]. MetaHuman Creator the Starting point of the Metaverse by Zhixin Fang, Li Bai Cai and Gang Wang. DOI: 10.1109/ISCTIS51085.2021.00040

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Editing. Exe Files

Sathish Kotturi[1], Shushank Balaji Reddy[1], Shibi Stephen K[1], Soumyadip Roy[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper has been undertaken to gain insights on how to edit .exe (Portable Executable) files and the methodologies used in it. The main purpose of creating this Research paper is to learn and understand from the perspective of a user who had written a complex program a long time ago and now wants to modify a small portion of it. Unfortunately, the original source code used to generate the executable file is lost and only the executable file is available. Under such a scenario the user can attempt to modify the executable file rather than rewriting the entire source code again with the required changes. A few tools are used in this paper to analyse the said executable file and make the required modifications to it. Depending on the requirement a few values can be changed entire sections of code can be appended or removed. This is done by modifying the binary of course that are present inside the executable file. The readjusting the values of jump instructions can prove to be difficult due to its underlying nature which relies on the position of the instruction in a given code. We are also able to modify the executable files provided by a third party even though we do not have the access to the original source code in which it was written. The resultant program can still be made secure against unauthorized modifications aimed at generating malicious software with the help of signed certificate [1]. Taking it a step further we can also use the provided debugger tools to analyze the behavior of the executable file and make optimizations to the machine code that would otherwise not be possible with high level language.

Keywords: Executable File, Portable Executable, Interactive Disassembly.

I. INTRODUCTION

An Exe file is an executable file which is derived after compiling a program that contains set of instructions. The exe file contains data that the Windows uses to read and run the program [2]. This data is saved in binary or compiled format which is often referred to as machine code. The .exe files are not like text files; we cannot edit it directly. Hence, we cannot get back to human readable source code as how it was written. However, with the help of certain tools like Interactive disassembler and python, we can certainly know the code in assembly language and edit them accordingly with the help of Python. This process is also called as reverse compilation. Hence, new functionalities can be incorporated in the .exe file once the source code is known.

A program is a sequence of instructions when compiled gives an .exe executable file. This file is generally not editable unlike any other files. Since the

format is not in human readable, it can get tricky to edit it. Let's say in real time, a program is written and compiled. The output can be verified and checked by running the .exe file. But if the source file is missing and we are left out with the .exe file, we can edit the if any changes are to be made and it will reflect accordingly. However, it is only possible using the following tools: Interactive disassembler – used to display the contents of the .exe file in assembly language.

Python – used to edit/make changes in the .exe file.

For example – A program to add two numbers is coded. 10+30=40.

The .exe file contains the output as 40. Using the above tools, we can successfully edit the .exe file and make changes such as 10+40=50.

Then the .exe will return the output as 50. Hence, with the help of certain tools like Interactive disassembler and python, we can certainly know the code in assembly language and edit them accordingly with the help of Python. This process is also called as reverse compilation. Hence, new functionalities can be incorporated in the .exe file once the source code is known.

II. METHODS AND MATERIAL

Interactive Disassembler

The Interactive Disassembler (IDA) is a disassembler for program which produces low level computing construct source code from machine-executable code. It upholds an assortment of executable organizations for various processors and working frameworks. It additionally can be utilized as a debugger for Windows PE, Mac OS X Mach-O, and Linux ELF executables. A decompiler module for programs incorporated with a C/C++ compiler is accessible at additional expense. The most recent full form of IDA Pro is business, while a prior and less proficient rendition is accessible for download for nothing (variant 7.6 as of March 2021)

IDA performs programmed code investigation, utilizing cross-references between code segments, information on boundaries of API calls, and other data. In any case, the idea of dismantling blocks complete exactness, and a lot of human mediation is fundamentally required; IDA has intelligent usefulness to support working on the dismantling [3]. An average IDA client will start with a naturally created dismantling posting and afterward convert segments from code to information as well as the other way around, rename, comment on, and in any case add data to the posting, until it turns out to be clear what it does.

Python files

Writing to a binary files - The open() function opens a document in text design of course. To open a record in double arrangement, add 'b' to the mode boundary. Subsequently the "rb" mode opens the record in parallel configuration for perusing, while the "wb" mode opens the document in double organization for composing. Dissimilar to message records, twofold documents are not intelligible. While opened utilizing any word processor, the information is unrecognizable. Reading to a binary files - Use the 'rb' mode in the open() function to read a binary files.

Python bytes

Python upholds a scope of types to store successions. There are six arrangement types: strings, byte successions (bytes objects), byte exhibits (bytearray objects), records, tuples, and range objects.

Strings contain Unicode characters. Their literals are written in single or twofold statements: 'python', "information". Bytes and bytearray objects contain single bytes - the previous is permanent while the last option is a variable grouping. Bytes array can be

developed the constructor, bytes(), and from literals; utilize a b prefix with typical string linguistic structure: b'python'. To build byte exhibits, utilize the byte array() function

bytes() and bytearray() functions

bytes() function: Return a new "bytes" object, which is an unchanging succession of little whole numbers in the reach 0 <= x < 256, print as ASCII characters when shown. bytes is an unchanging form of bytearray - it has the equivalent non-changing techniques and a similar ordering and cutting way of behaving.

bytearray() function : Return another variety of bytes. The bytearray type is an impermanent succession of whole numbers in the reach 0 <= x < 256. It has a large portion of the standard strategies for variable groupings, portrayed in Mutable Sequence Types, as well as most techniques that the bytes type has, see Bytes and Byte Array Methods.

Ctypes in python

Ctypes is an foreign function library for Python. It gives C compatible data types, and permits calling capacities in DLLs or shared libraries. Enveloping these libraries by unadulterated Python can be utilized. ctypes sends out the cdll, and on Windows likewise windll and oledll objects to stack dynamic connection libraries.

We load libraries by getting to them as traits of these articles. cdll loads libraries which commodity functions utilizing the standard cdecl calling show, while windll libraries call functions utilizing the stdcall calling show. oledll likewise utilizes the stdcall calling show, and accepts the function return a Windows HRESULT blunder code. The mistake code is utilized to naturally raise WindowsError Python exemptions when the function call falls flat.

III. RESULTS AND DISCUSSION

The structure of a portable executable file is shown in the figure. As mentioned before exe files are just portable executable files containing various sections of machine level code or opcodes. The header of a portable executable file contains three major sections which are coff header, optional header and section table. The coff header contains information about various aspects of the executable file, like the architecture of the machine it was designed for, number of sections that it contains, size of the optional header etc. The optional header whose name is misleading is not optional at all and it contains crucial information like the size of code, initialized data and uninitialized data address of entry point base of code base of data and other various addresses and sizes. The section table contains the information about all the sections that are present in the executable file [4]. In practice it is possible to edit and manually restructure all the sections of the portable executable file but it is beyond the scope of this paper. In this paper we majorly focus on editing the content of the .text section.

The Original Executable File

Let us consider a simple executable file which was generated on compiling a C Program on a windows machine. The program contains two inbuilt integers which it adds and displays the output using the printf function. As a demonstration we shall attempt to modify the said exe file and change one of its integers that is adding to obtain the result. The output produced when executing the original executable file is shown in the figure below.

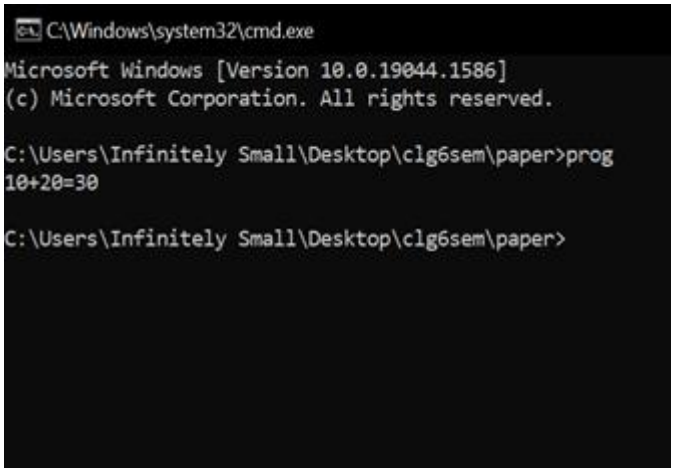


Fig. 1: Original Program

Interactive Disassembler

The Interactive Disassembler software is first used to analyse the given exe file and display various parts of its contents. The disassembler also shows various bytes that are present in the exe file, their location, and their respective function [5]. Though in principle, Interactive Disassembler alone provides the functionality to edit the said executable file, it is only available in the paid version which has to be purchased on the basis of an expiring licence [6]. Hence, we chose an alternative method for performing the edit on the said file.

```
push    rbp
mov     rbp, rsp
sub     rsp, 30h
call    __main
mov     [rbp+var_4], 0Ah
mov     [rbp+var_8], 14h
mov     edx, [rbp+var_4]
mov     eax, [rbp+var_8]
add     eax, edx
mov     [rbp+var_C], eax
mov     ecx, [rbp+var_C]
mov     edx, [rbp+var_8]
mov     eax, [rbp+var_4]
mov     r9d, ecx
mov     r8d, edx
mov     edx, eax
lea     rcx, Format     ; "%d+%d=%d\n"
call    printf
nop
add     rsp, 30h
pop     rbp
retn
main endp
```

Fig. 2: Interactive Disassembler View of Prog

Nonetheless Interactive Disassembler even with its limited functionality in the free version played a crucial role in simplifying the task of having to parse the various section headers, go to the required section, and scan through the opcodes until the desired opcode is found. Which would have certainly been a humongous task on its own.

```
48 8D 0D 17 00 00   00 48 83 C4 20 5D
00 48 8D 05 96 FF   FF FF EB DD 0F 1F
89 E5 5D C3 90 90   90 90 90 90 90 90
89 E5 48 83 EC 30   E8 B3 01 00 00 C7
00 00 C7 45 F8 14   00 00 00 8B 55 FC
D0 89 45 F4 8B 4D   F4 8B 55 F8 8B 45
41 89 D0 89 C2 48   8D 0D 42 2A 00 00
00 90 48 83 C4 30   5D C3 90 90 90 90
83 EC 30 48 89 CB   48 8B 0D 01 64 00
00 00 48 83 F8 FF   48 89 44 24 20 74
00 00 E8 AF 15 00   00 48 8B 0D E0 63
```

Fig. 3: Interactive Disassembler HEX View

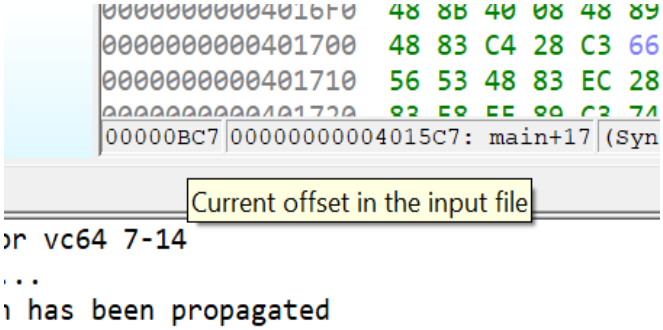


Fig. 4: Current offset in the input file

Python Code

As mentioned above, for better ease of access we decided to use an alternative approach for editing the executable file. An exe file, at the end of the day is still a binary file stored on the hard drive so in principle any programming language should have the capability necessary to edit such a file. We decided to use Python programming language to edit the given exe file and modify the required byte at the specified position to change the integer 20 with the hex code 0x14 to 30 with the hex code 0x1e which is shown in the figure. Python is an easy to use yet a powerful programming language which as demonstrated can edit the executable file and make the necessary changes that are required by the application

programmer. Python provides built in functionality to read the binary file with the open function and specifying the parameter as 'rb' which stands for read binary. Following which the f.read function reads the file into a Python bytes object. Unfortunately, the bites object of python is an immutable object and hence its contents cannot be altered [7]. Which leaves us with having to slice the bytes object and then reconcatinate the slices into a new bytes object. After this we again use the open function with 'wb' parameter which stands for write binary to create a new binary file and the subsequent f.write function writes the bytes object into the binary file [8].

editEXE.py

```
1  with open('prog.exe','rb') as f:
2      a=f.read()
3
4  a2=a[:0xbc7]+b'\x1e'+a[0xbc8:]
5
6  with open('prog2.exe','wb') as f:
7      f.write(a2)
8
```

Fig. 5: Python code to change the required byte.

Modified Executable File

As shown in the figure the newly created executable file is Run and its output has the changed value of 10 + 30 = 40 which serves as a proof of concept for the agenda being discussed in this paper.

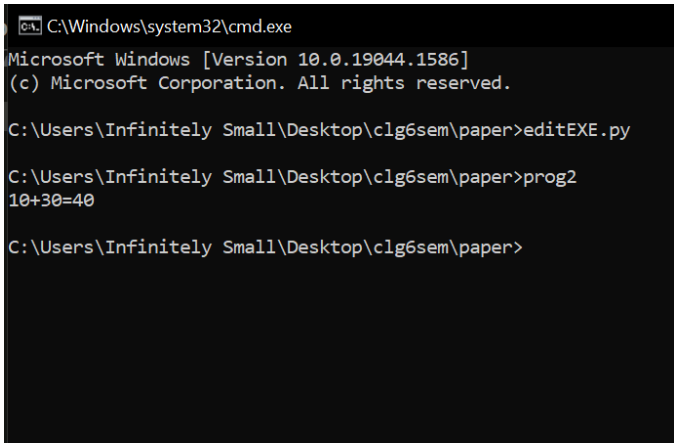


Fig. 6: Output of the modified exe file

IV. CONCLUSION

As demonstrated above we have shown that a portable executable file or exe file can be successfully edited using the interactive disassembler software and Python programming language. The interactive disassembler is successfully used to analyse the executable file and locate the byte which needs to be changed in order to obtain the desired result, whereas the Python programming language is used to perform the actual edit where the file is opened and read from, the necessary changes are made to the read data after which the changed data is written on to a new file. As this is only a demonstration, more complicated edits can also be performed on the executable file where rather than changing a single byte, an entire chunk of code can be either appended or inserted into the executable file. In principle this technique can also be used to inject malicious code into executable files which are commonly used in popular software which is prevented by using signing certificate which cannot be forged without the private key of the original signer. Hence this method can be safely used to edit the executable file as the programmer desires.

V. REFERENCES

[1]. File Identification and Profiling by James M. Aquilina, in Malware Forensics, 2008
[2]. https://en.wikipedia.org/wiki/Portable_Executable
[3]. https://hex-rays.com/ida-free/
[4]. https://docs.microsoft.com/en-us/windows/win32/debug/pe-format
[5]. https://en.wikipedia.org/wiki/Interactive_Disassembler
[6]. https://hex-rays.com/ida-pro/
[7]. https://docs.python.org/3/library/stdtypes.html#bytes-objects
[8]. https://docs.python.org/3/tutorial/inputoutput.html

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Analysis on Cost-cutting with Virtualization

Jashwanth M S[1], Kasish S V[1], UmaShankar Reddy[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper presents a brief look at virtualization, and its benefits in today's world. Virtualization results in overall reduction in infrastructure and operating costs, manpower requirements and space. Virtualization provides benefits such as greater efficiency in CPU utilization, better management through central environment control, reduced project timelines, improved disaster recovery capability, more central control of the desktop, and also improved outsourcing services. It discusses the reduction of hardware Costs incurred by small-scale industries and Educational Institutions.

Keywords: Virtualization, Cost reduction, Efficiency, Better management

I. INTRODUCTION

Virtualization is becoming a core business strategy from desktop to data centre. It can provide an isolated execution environment to applications, support share and reuse of hardware resources. The cost reduction and simplicity of management are the prominent advantages. It can help to deliver high priority business services more quickly. Enhance energy efficiency. Optimizing the various resources with virtualization helps to improve organization efficiency. Here, the various optimization methods like desktop, server, network and storage are included. Optimization is the process of making something more efficient or optimal. Optimization is about virtualization of storage, servers and move towards on demand computing. Desktop optimization enables organizations to centrally define virtual machines (VMs) and assign these VMs to authenticated users to run on their PCs[4]. It offers temporary access to a corporate desktop, and provides higher security for corporate applications and data.

It is noted that majority of the system in small scale industries or educational Institutions waste a lot of money in terms of Maintenance and hardware costs to provide individual systems to everyone who would use their services, this cost can be cut down by Virtualization as the VMs are maintained by the service providers and the infrastructural cost reduces significantly[1].

II. LITERATURE

Gitanjali Gupta has listed the benefits availed by Server Virtualization [11] where virtual desktop infrastructure and analysis on Citrix is done citing the advantages of Citrix over Windows Server.

Aaqib Rashid and Amit Chaturvedi [13] have discussed virtualization with cloud computing and its role in today's architecture.

Dario Trabucco [15] has stated the advantages of using virtualization in small-scale organizations with accurate details in his own organization

NazihSalhab , Rana Rahim , and Rami Langa [12] have discussed the relevancy of Virtualization and Cloud Computing with the rise of 5G services. The emphasis is laid on reducing costs related to Storage and Memory.

III. RESULTS AND DISCUSSION

A. TYPES OF VIRTUALIZATIONS

1) OS virtualization:Virtualizing an operating system environment is the most common form of virtualization. It involves putting a second instance or multiple instances of an operating system, like Windows, on a single machine. This empowers businesses to reduce the amount of physical hardware required to run their software by cutting down the number of actual machines. It saves companies cash on energy, cabling, hardware, rack space, and more, while still allowing them to run the same quantity of applications[14].
2) Application-Server Virtualization: This is another large presence in the virtualization space, and has been around since the inception of the concept. It is often referred to as 'advanced load balancing', as it spreads applications across servers, and servers across applications. This enables IT departments to balance the workload of specific software in an agile way that doesn't overload a specific server or underload a specific application in the event of a large project or change. In addition to load balancing, it also allows for easier management of servers and applications, since you can manage them as a single instance[14]. Additionally, it gives way to greater network security, as only one server is visible to the public while the rest are hidden behind a reverse proxy network security appliance.
3) Hardware Virtualization Hardware virtualization is one of the rarer forms of virtualization, and when simply explained it is similar to OS virtualization (it is, in fact, often required for OS virtualization). Except, instead of putting multiple software instances on a single machine, chunks of a machine are partitioned off to perform specific tasks [12].
4) Storage Virtualization: Storage virtualization is an array of servers that are managed by a virtual storage system. The servers aren't aware of exactly where their data is, and instead function more like worker bees in a hive.
5) Administrative Virtualization: Administrative virtualization is one of the least-known forms of virtualization likely due to the fact that it's primarily used in data centers. The concept of administration, or 'management,' virtualization means segmented admin roles through group and user policies. For example, certain groups may have access to read specific servers, infrastructure, application files, and rules, but not to change them[2].

B. BENEFITS OF VIRTUALIZATION

1) Slash your IT expenses: Utilizing a non-virtualized environment can be inefficient because when you are not consuming the application on the server, the compute is sitting idle and can't be used for other applications. When you virtualize an environment, that single physical server transforms into many virtual machines. These virtual machines can have different operating systems and run different applications while still all being hosted on a single physical server [10]. The consolidation of the applications onto virtualized environments is a

more cost-effective approach because you'll be able to consume fewer physical customers, helping you spend significantly less money on servers and bring cost savings to your organization.

2) Reduce downtime and enhance resiliency in disaster recovery situations: When a disaster affects a physical server, someone is responsible for replacing or fixing it—this could take hours or even days. With a virtualized environment, it's easy to provision and deploy, allowing you to replicate or clone the virtual machine that's been affected. The recovery process would take mere minutes—as opposed to the hours it would take to provision and set up a new physical server—significantly enhancing the resiliency of the environment and improving business continuity.
3) Increase efficiency and productivity: With fewer servers, your IT teams will be able to spend less time maintaining the physical hardware and IT infrastructure [11]. You'll be able to install, update, and maintain the environment across all the VMs in the virtual environment on the server instead of going through the laborious and tedious process of applying the updates server-by-server. Less time dedicated to maintaining the environment increases your team's efficiency and productivity.
4) Control independence and DevOps: Since the virtualized environment is segmented into virtual machines, your developers can quickly spin up a virtual machine without impacting a production environment. This is ideal for Dev/Test, as the developer can quickly clone the virtual machine and run a test on the environment [10]. For example, if a new software patch has been released, someone can clone the virtual machine and apply the latest software update, test the environment, and then pull it into their production application. This increases the speed and agility of an application
5) Move to be more green-friendly (organizational and environmental): When you are able to cut down on the number of physical servers you're using, it'll lead to a reduction in the amount of power being consumed [12]. This has two green benefits: It reduces expenses for the business, and that money can be reinvested elsewhere. It reduces the carbon footprint of the data center.

C. INEFFICIENT USAGE OF RESOURCES IN TRADITIONAL COMPUTING

With Traditional Computing, organizations used to purchase hardware keeping future requirements in mind such as projected increase in server load or memory required. This caused lot of resources to be wasted knowingly. Sometimes this wastage of resources would be at the expense of some other important project because upfront huge investment was done on high end hardware. Dynamic allocation of resources because of server virtualization helps in optimal utilization of hardware. Additional CPU or memory can be allocated when required thus saving the wastage and reducing the upfront investment. Not only this, additional applications can also be supported on the same hardware.

For instance, an Educational Institution has several machines in labs and libraries. Each physical machine consumes a lot of power, cooling & space. All these costs are incurred by the Institution when most of the machines aren't used to its full capacity [8]. Virtualization enables you to add more virtual servers without incurring additional costs on these parameters. In addition to this, it also reduces cost by requiring few Keyboard-Video-Mouse systems, backup tapes, power cables and network cables to connect devices.

Security Issues with Traditional Computing in Institutions is also a major concern, as all the machines are publicly available and a student's work/project can be accessed by probably malicious users within the same institution. This can lead to Tampering, copying or even deleting those files.

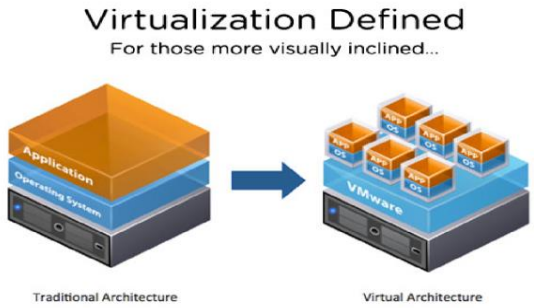


Fig 1: Graphical representation of Virtualization [1]

D. SAVING RESOURCES WITH VIRTUALIZATION

1) Reducing the Overhead of Physical Servers: When an application is virtualized, it runs on a shared server equipped with the appropriate hypervisor or virtual machine management firmware [10]. With this capability, the need for a dedicated server and any other ancillary support equipment for a given application is eliminated. This not only reduces the initial cost of hardware acquisition and installation, but also the ongoing cost of time and maintenance of the IT team. Reduction in server numbers also increases availability of floor space, which can be utilized for other purposes.
2) Software and Licensing Costs: The cost for software licensing is another area where expenses are reduced in a virtual system. Whereas the standard model was one license for each server, licensing models with virtual applications have become more diverse and cheaper. With multiple virtual servers running on a single physical server, it is impractical to require a 1 to 1 license. For example, Cisco states that they allow over 150 virtual server licenses per physical server, with unlimited operating system guest licenses [13]. This new system allows customers to save up to 80% in licensing expenses.
3) Electrical Costs: This feature may not appear to be a large component of cost savings, but electrical costs factor into the entire package. A study [15] in 2012 from the Uptime Institute found that decommissioning a single 1U rack server can save $500 annually in energy. VMware found that server consolidation can reduce energy cost up to 80%.
4) Decreased Time for Server Provisioning and Deployment: With a virtualized server, system provisioning and deployment can be done in a fraction of the time compared to setting up a new physical server. Not only is the time spent requisitioning the server, obtaining approvals, and sending out a PO eliminated, but the hours it takes to install the operating system and application are reduced to minutes when cloning an existing virtual machine. All this time savings can add up to a substantial savings in labour costs.

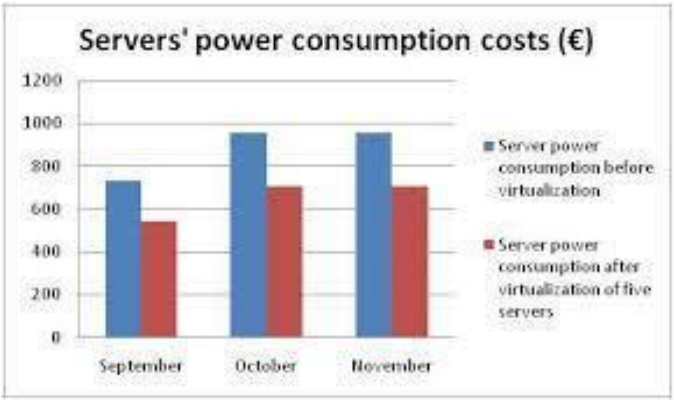


Fig 2: Comparison with Traditional Computing

E. CHALLENGES IN VIRTUALIZATION

1) Resource distribution: The way virtualization partitions systems can result in varied ways —

some might function really well, and others might not provide users access to enough resources to meet their needs. Resource distribution problems often occur in the shift to virtualization and can be fixed by working on capacity planning with your service provider.

2) VM Sprawl: VM sprawl, the unchecked growth of virtual machines in a virtual environment, as any virtualization admin knows, can cripple an otherwise healthy environment. It is problematic because its underlying cause often stays hidden until it manifests in resource shortages. You should look at how virtual machines will be managed, who will be doing what, and what systems you're going to use. One of the optimal times to develop an overall management plan is when you're in a testing phase, before migration.
3) Backward compatibility: Using legacy systems can cause problems with newer virtualized software programs. Compatibility issues can be time-consuming and difficult to solve. A good provider may be able to suggest upgrades and workarounds to ensure that everything functions the way they should.
4) Performance monitoring: Virtualized systems don't lend themselves to the same kind of performance monitoring as hardware like mainframes and hardware drives do. Try tools like VMmark to create benchmarks that measure performance on virtual networks and to monitor resource usage as well.

IV. CONCLUSION

Virtualization allows maximum utilization of resources, as result of this we can use minimum resources for doing maximum computation. Virtualization in cloud provides an easy way to set up new virtual servers, so you don't have to manage a lot of them. Keeping track of your physical resources that are used for virtual resources is vital, so shop for solutions that have easy-to-use tools that help you measure and monitor usage.

Each physical machine consumes a lot of power, cooling & space. Virtualization enables you to add more virtual servers without incurring additional costs on these parameters. At the same time there can be drawbacks to this approach too, Failure of the central server or node would lead to failure of the whole network or would certainly affect the coordination of the nodes in that network. Thus taking into consideration of the drawbacks that are mentioned, the customer must fix a definite SLA with the Service provider with the help of a moderator who helps in monitoring the services and negotiates the violations. Virtualization doesn't turn out to be good for everyone, But in most cases, the efficiency, productivity, security and cost advantages outweigh any issues, and thus, virtualization is continuously gaining popularity.

V. REFERENCES

[1]. https://www.higherground.com/post/2018/09/19/virtualization-reducing-costs
[2]. Research on Computing Resources Sharing of University Laboratories in Education Cloud Environment, IOP Conference Series Materials Science and Engineering
[3]. https://www.javatpoint.com/virtualization-in-cloud-computing
[4]. http://www.itbriefcase.net/network-server-storage-and-desktop-virtualization
[5]. Book-Mastering Cloud-computing ,https://www.geeksforgeeks.org/virtualization-a-machine-reference-model/
[6]. Alex Fishman; Mike Rapoport; EvgenyBudilovsky; IzikEidus (2013-06-25).

"HVX: Virtualizing the Cloud" (PDF). rackcdn.com. Retrieved 2014-12-16.

[7]. S. Perez, —Mobile cloud computing: $9.5 billion by 2014‖, http://exoplanet.eu/catalog.php, 2010.

[8]. "Virtualization in education" (PDF). IBM. October 2007. Retrieved 6 July 2010. A virtual computer is a logical representation of a computer in software.

[9]. Orit Wasserman, Red Hat (2013). "Nested virtualization: Shadow turtles" (PDF). KVM forum. Retrieved 2021-05-07.

[10]."Desktop Virtualization Tries to Find Its Place in the Enterprise". Dell.com. Retrieved 2012-06-19.

[11].Reduction in Infrastructure and operating costs using Server Virtualization, International Journal Of Engineering And Computer Science ISSN: 2319-7242

[12].Virtualization and its Role in Cloud Computing Environment, E-ISSN: 2347-2693

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Case Study on Neural Networks

Abhishek. L[1], Mrs. Champavathy. B[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The term "Neural networks" is a very evocative one. It suggests machines that are something like brains and is potentially laden with the science fiction connotations of the Frankenstein mythos. One of the main tasks of this book is to demystify neural networks and show how, while they indeed have something to do with brains, their study also makes contact with other branches of science, engineering and mathematics. A neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the interunit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns. Nevertheless, all symbols and expressions will be explained as they arise so that, hopefully, these should not get in the way of the essentials: that is, concepts and ideas that may be described in words. This chapter is intended for orientation. We attempt to give simple descriptions of what networks are and why we might study them. In this way, we have something in mind right from the start, although the whole of this book is, of course, devoted to answering these questions in full.

I. INTRODUCTION

To understand the neural networks, firstly the functions of actual neurons must be discussed. Neurons (also called neurons or nerve cells) are the fundamental units of the brain and nervous system, the cells responsible for receiving sensory input from the external world, for sending motor commands to our muscles, and for transforming and relaying the electrical signals at every step in between. More than that, their interactions define who we are as people. Having said that, our roughly 100 billion neurons do interact closely with other cell types, broadly classified as glia (these may actually out number neurons, although it's not really known).for the last two decades. [3] A useful analogy is to think of a neuron as a tree. A neuron has three main parts: dendrites, an axon, and a cell body or soma, which can be represented as the branches, roots and trunk of a tree, respectively. A dendrite (tree branch) is where a neuron receives input from other cells. Dendrites branch as they move towards their tips, just like tree branches do, and they even have leaf-like structures on them called spines. The axon (tree roots) is the output structure of the neuron; when a neuron wants to talk to another neuron, it sends an electrical message called an actionpotential throughout the entire axon. The soma (tree trunk) is where the nucleus lies, where the neuron's DNA is housed, and

where proteins are made to be transported throughout the axon and dendrites.

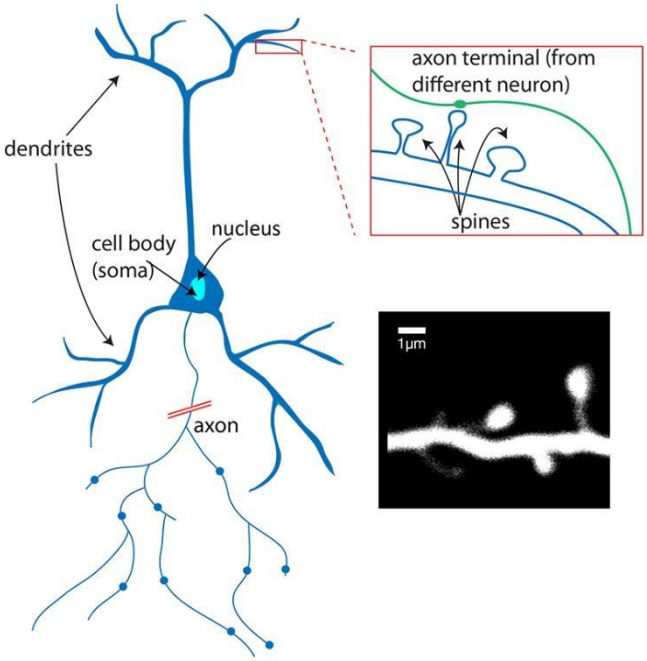


FIG. 1: GRAPHICAL REPRESENTATION OF A BIOLOGICAL NEURON

[3] The tree-like structure of a neuron. Dendritic spines are small structures that receive inputs from the axons of other neurons. Bottom-right image: a segment of dendrite from which spines branch off, like leaves off a tree branch. Note the very small size (~0.001mm). (Image: Alan Woodruff; De Roo et al / CC BY-SA 3.0 via Commons)

II. CONCEPTS AND DEFINITIONS

Axon – The long, thin structure in which action potentials are generated; the transmitting part of the neuron. After initiation, action potentials travel down axons to cause release of neurotransmitter.

Dendrite – The receiving part of the neuron. Dendrites receive synaptic inputs from axons, with the sum total of dendritic inputs determining whether the neuron will fire an action potential.Spine – The small protrusions found on dendrites that are, for many synapses, the postsynaptic contact site.Action potential – Brief electrical event typically generated in the axon that signals the neuron as 'active'. An action potential travels the length of the axon and causes release of neurotransmitter into the synapse. The action potential and consequent transmitter release allow the neuron to communicate with other neurons.[2]Artificial neural networks (ANNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

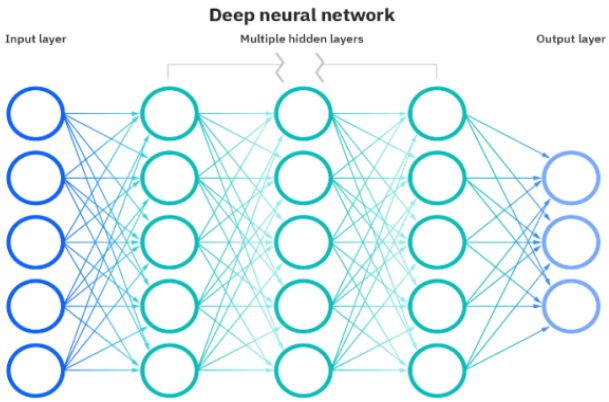


FIG. 2: DEEP NEURAL NETWORK

[4] Neural networks rely on training data to learn and improve their accuracy over time. However, once these learning algorithms are fine-tuned for accuracy, they are powerful tools in computer science and artificial intelligence, allowing us to classify and cluster data at a high velocity. Tasks in speech recognition or image recognition can take minutes versus hours when compared to the manual identification by human experts. One of the most well-known neural networks is Google's search algorithm.

WORKING OF THE NEURAL NETWORK

[4]Think of each individual node as its own linear regression model, composed of input data, weights, a bias (or threshold), and an output. The formula would look something like this:

$$\sum_{i=1}^{m} w_i x_i + bias = w_1 x_1 + w_2 x_2 + w_3 x_3 + bias$$

∑wixi + bias = w1x1 + w2x2 + w3x3 + bias

$$output = f(x) = \begin{cases} 1 \text{ if } \sum w_1 x_1 + b \geq 0 \\ 0 \text{ if } \sum w_1 x_1 + b < 0 \end{cases}$$

output = f(x) = 1 if ∑w1x1 + b>= 0; 0 if ∑w1x1 + b < 0

Once an input layer is determined, weights are assigned. These weights help determine the importance of any given variable, with larger ones contributing more significantly to the output compared to other inputs. All inputs are then multiplied by their respective weights and then summed. Afterward, the output is passed through an activation function, which determines the output. If that output exceeds a given threshold, it "fires" (or activates) the node, passing data to the next layer in the network. This results in the output of one node becoming in the input of the next node. This process of passing data from one layer to the next layer defines this neural network as a feedforward network.[3]Let's break down what one single node might look like using binary values. We can apply this concept to a more tangible example, like whether you should go surfing (Yes: 1, No: 0). The decision to go or not to go is our predicted outcome, or y-hat. Let's assume that there are three factors influencing your decision-making:

1. Are the waves good? (Yes: 1, No: 0)
2. Is the line-up empty? (Yes: 1, No: 0)
3. Has there been a recent shark attack? (Yes: 0, No: 1) Then, let's assume the following, giving us the following inputs:

- X1 = 1, since the waves are pumping
- X2 = 0, since the crowds are out
- X3 = 1, since there hasn't been a recent shark attack Now, we need to assign some weights to determine importance. Larger weights signify that particular variables are of greater importance to the decision or outcome.
- W1 = 5, since large swells don't come around often
- W2 = 2, since you're used to the crowds
- W3 = 4, since you have a fear of sharksFinally, we'll also assume a threshold value of 3, which would translate to a bias value of –3. With all the various inputs, we can start to plug in values into the formula to get the desired output.

Y-hat = (1*5) + (0*2) + (1*4) – 3 = 6

[5]If we use the activation function from the beginning of this section, we can determine that the output of this node would be 1, since 6 is greater than 0. In this instance, you would go surfing; but if we adjust the weights or the threshold, we can achieve different outcomes from the model. When we observe one decision, like in the above example, we can see how a neural network could make increasingly complex decisions depending on the output of previous decisions or layers.In the example above, we used perceptrons to illustrate some of the mathematics at play here, but neural networks leverage sigmoid neurons, which are distinguished by having values between 0 and 1. Since neural networks behave similarly to decision trees, cascading data from one node to another, having x values between 0 and 1 will reduce the impact of any given change of a single variable on the output of any given node, and subsequently, the output of the neural network.[5]As we start to think about more practical use cases for neural networks, like image recognition or classification, we'll leverage supervised learning, or labeled datasets, to train the algorithm. As we train

the model, we'll want to evaluate its accuracy using a cost (or loss) function. This is also commonly referred to as the mean squared error (MSE). In the equation below,

- *i* represents the index of the sample,
- y-hat is the predicted outcome,
- y is the actual value, and *m* is the number of samples

$$Cost\ Function = MSE = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

CostFunction= *MSE*=1/2*m*
∑129_(*i*=1)^*m*▒(*y*^((*i*))–*y*^((*i*)))^2

Ultimately, the goal is to minimize our cost function to ensure correctness of fit for any given observation. As the model adjusts its weights and bias, it uses the cost function and reinforcement learning to reach the point of convergence, or the local minimum. The process in which the algorithm adjusts its weights is through gradient descent, allowing the model to determine the direction to take to reduce errors (or minimize the cost function). With each training example, the parameters of the model adjust to gradually converge at the minimum.

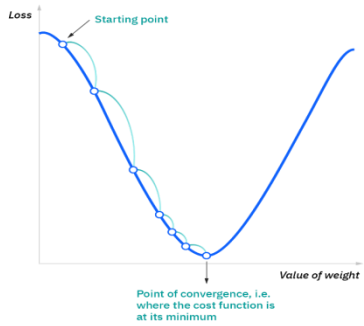


Fig. 3: Graphical Representation of Reinforcement learning and Cost function

III. TYPES OF NEURAL NETWORKS

[6]Neural networks can be classified into different types, which are used for different purposes. While this isn't a comprehensive list of types, the below would be representative of the most common types of neural networks that you'll come across for its common use cases:The perceptron is the oldest neural network, created by Frank Rosenblatt in 1958. It has a single neuron and is the simplest form of a neural network

[5]Feedforward Neural Network – Artificial Neuron.[5]This neural network is one of the simplest forms of ANN, where the data or the input travels in one direction. The data passes through the input nodes and exit on the output nodes. This neural network may or may not have the hidden layers. In simple words, it has a front propagated wave and no backpropagation by using a classifying activation function usually.Below is a Single layer feed-forward network. Here, the sum of the products of inputs and weights are calculated and fed to the output. The output is considered if it is above a certain value i.e threshold(usually 0) and the neuron fires with an activated output (usually 1) and if it does not fire, the deactivated value is emitted (usually -1).

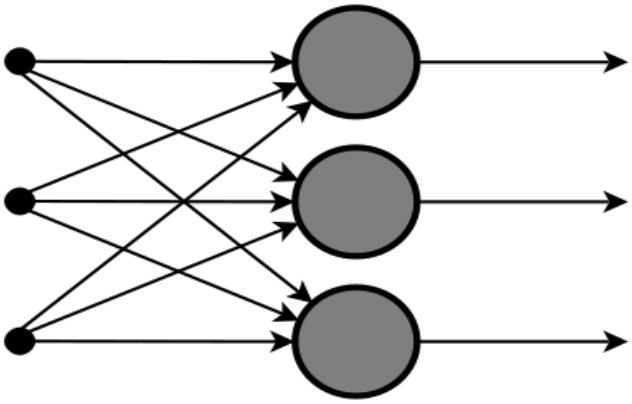


FIG. 4: FEEDFORWARD NEURAL NETWORK

Radial basis function Neural Network:

[5] Radial basic functions consider the distance of a point with respect to the center. RBF functions have two layers, first where the features are combined with the Radial Basis Function in the inner layer and then the output of these features are taken into

consideration while computing the same output in the next time-step which is basically a memory.Below is a diagram that represents the distance calculating from the center to a point in the plane similar to a radius of the circle. Here, the distance measure used in euclidean, other distance measures can also be used. The model depends on the maximum reach or the radius of the circle in classifying the points into different categories. If the point is in or around the radius, the likelihood of the new point begin classified into that class is high. There can be a transition while changing from one region to another and this can be controlled by the beta function.

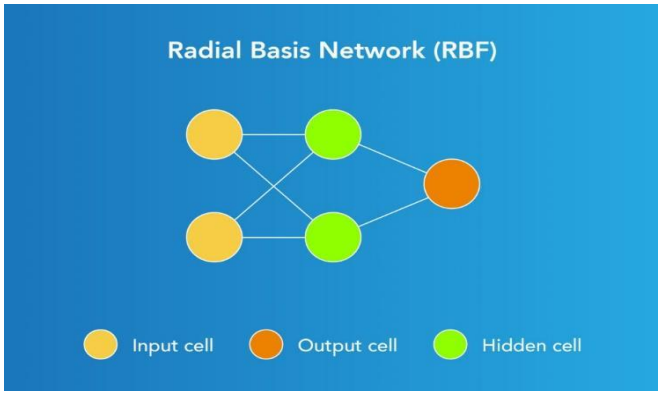


FIG. 4: Radical Basis Neural Network

Recurrent Neural Networks

Designed to save the output of a layer, Recurrent Neural Network is fed back to the input to help in predicting the outcome of the layer. The first layer is typically a feed forward neural network followed by recurrent neural network layer where some information it had in the previous time-step is remembered by a memory function. Forward propagation is implemented in this case. It stores information required for it's future use. If the prediction is wrong, the learning rate is employed to make small changes. Hence, making it gradually increase towards making the right prediction during the backpropagation.

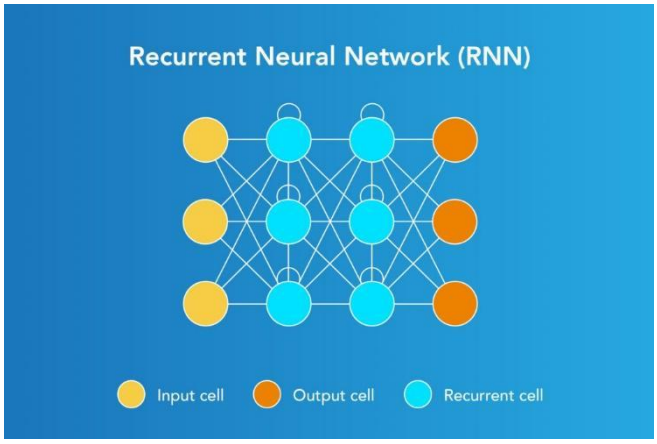


FIG. 5: Recrrent Neural Network

IV. ADVANTAGES OF NEURAL NETWORK

[4]ANN outputs aren't limited entirely by inputs and results given to them initially by an expert system. This ability comes in handy for robotics and pattern recognition systems. This neural network has the potential for high fault tolerance and can debug or diagnose a network on its own. ANN can go through thousands of log files from a company and sort them out. It is currently a tedious task done by administrators, but it will save a significant amount of time, energy, and resources if it can be automated. Nonlinear systems can find shortcuts to reach computationally expensive solutions. We see this in the banking industry, for example, where they work on a particular Excel spreadsheet, and as time goes by, start building codes around it. In over 20 years, they might create a repertoire of all these functions, and the neural network rapidly comes up with the same answers otherwise done in days, weeks, or even a month, when done by a large bank.

[7]The architecture of a convolutional neural network is a multi-layered feed-forward neural network, made by stacking many hidden layers on top of each other in sequence. It is this sequential design that allows convolutional neural networks to learn hierarchical features.The hidden layers are typically convolutional

layers followed by activation layers, some of them followed by pooling layers.A simple convolutional neural network that aids understanding of the core design principles is the early convolutional neural network LeNet-5, published by Yann LeCun in 1998. LeNet is capable of recognizing handwritten characters. The process of training a convolutional neural network is fundamentally the same as training any other feedforward neural network, and uses the back propagation algorithm.Initially, the network is created with random values in all of its weights and biases. The training examples consist of a set of tuples of images and classes. Let us imagine the case of training a convolutional neural network to categorize images as 'cat' or 'dog'.Each training image is passed through the entire network and the final SoftMax outputs a vector containing probability For example the output

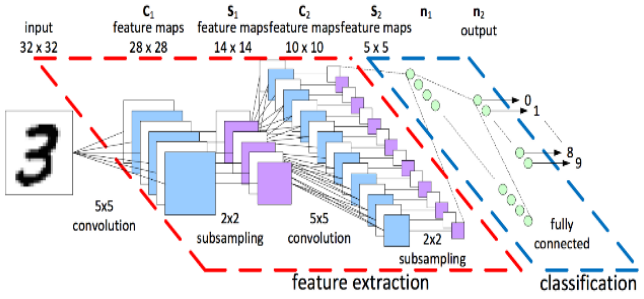


FIG. 5: Image classification using CNN

[8]The image classification pipeline. We've seen that the task in Image Classification is to take an array of pixels that represents a single image and assign a label to it. Our complete pipeline can be formalized as follows:

Input: Our input consists of a set of *N* images, each labeled with one of *K* different classes. We refer to this data as the training set.Learning: Our task is to use the training set tolearn what every one of the classes looks like. We refer to this step as training *a* classifier, or learning a model.

Evaluation: In the end, we evaluate the quality of the classifier by asking it to predict labels for a new set of images that it has never seen before.[8] We will then compare the true labels of these images to the ones predicted by the classifier. Intuitively, we're hoping that a lot of the predictions match up with the true answers (which we call the ground truth).As our first approach, we will develop what we call a Nearest Neighbor Classifier. This classifier has nothing to do with Convolutional Neural Networks and it is very rarely used in practice, but it will allow us to get an idea about the basic approach to an image classification problem.Example image classification dataset: CIFAR-10. One popular toy image classification dataset is the CIFAR-10 dataset. This dataset consists of 60,000 tiny images that are 32 pixels high and wide. Each image is labeled with one of 10 classes (for example "airplane, automobile, bird, etc"). These 60,000 images are partitioned into a training set of 50,000 images and a test set of 10,000 images. In the image below you can see 10 random example images from each one of the 10 classes:Suppose now that we are given the CIFAR-10 training set of 50,000 images (5,000 images for every one of the labels), and we wish to label the remaining 10,000. The nearest neighbor classifier will take a test image, compare it to every single one of the training images, and predict the label of the closest training image. In the image above and on the right you can see an example result of such a procedure for 10 example test images. Notice that in only about 3 out of 10 examples an image of the same class is retrieved, while in the other 7 examples this is not the case. For example, in the 8th row the nearest training image to the horse head is a red car, presumably due to the strong black background. As a result, this image of a horse would in this case be mislabeled as a car.You may have noticed that we left unspecified the details of exactly how we compare two images, which in this

case are just two blocks of 32 x 32 x 3. One of the simplest possibilities is to compare the images pixel by pixel and add up all the differences. In other words, given two images and representing them as vectors I1,I2I1,I2 , a reasonable choice for comparing them might be the L1 distance:

$d1(I1,I2)=\sum p|Ip1-Ip2|$

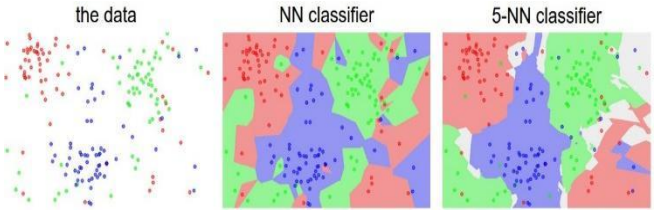


FIG. 6: K-nearest neighbour classifier

An example of the difference between Nearest Neighbor and a 5-Nearest Neighbor classifier, using 2-dimensional points and 3 classes (red, blue, green). The colored regions show the decision boundaries induced by the classifier with an L2 distance. The white regions show points that are ambiguously classified (i.e. class votes are tied for at least two classes). Notice that in the case of a NN classifier, outlier datapoints (e.g. green point in the middle of a cloud of blue points) create small islands of likely incorrect predictions, while the 5-NN classifier smooths over these irregularities, likely leading to better generalization on the test data (not shown). Also note that the gray regions in the 5-NN image are caused by ties in the votes among the nearest neighbors

V. CONCLUSION

With the rapid pace that AI and machine learning are being adopted by companies today, we could see more advancements in the applications of neural networks in the foreseeable future. AI and machine learning will offer a wealth of personalized choices for users worldwide. For example, all mobile and web applications try to give you an enhanced customized experience based on your search history, and neural networks can make that possible. Hyper-intelligent virtual assistants will make life easier. If you have ever used Google assistant, Siri, or any other products, you can see how they're slowly evolving. They may even predict your email responses in the future!

VI.REFERENCES

[1]. https://europepmc.org/article/med/19065803
[2]. https://qbi.uq.edu.au/brain/brainanatomy/what-neuron#:~:text=Neurons%20(also%20called%20neurones%20or,at%20every%20step%20in%20between.
[3]. Queens Brain Institute, Brain Anatomy Author: Dr. Alan Woodruff
[4]. https://www.ibm.com/cloud/learn/neural-networks
[5]. https://analyticsindiamag.com/6-types-of-artificial-neural-networks -currently-being-used-in-todays-technology/
[6]. https://www.mygreatlearning.com/blog/types-of-neural-networks/
[7]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-neural-network
[8]. Deep Learning Case Studies Neural Networks Author: Danial Graupe
[9]. https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network
[10].https://cs231n.github.io/classification/

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Role of Linux in Microsoft

Vijay Selvaraj J[1], Prakruthi C Prakash[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Microsoft has been one of the most used operating systems world-wide. Being an open-source platform, Microsoft has given its users one of the best user interfaces and functionalities focused mainly on graphical user interface. Though it is the most used and highly recommended operating systems, the operating system that Microsoft uses in its workspace and servers is not Windows but Linux. Why couldn't Microsoft make an operating system of its own to run its servers. To understand this, we have to first understand the features of Linux, how Linux servers work, its advantages and the drawbacks of Microsoft in detail.

Keywords: —Operating System, Windows, Linux, Open source.

I. INTRODUCTION

As we all know, Linux[1] is a free and open-source operating system and its source code is available for all users to modify and use for commercial and non-commercial uses. It was initially designed for personal computers, but over the time, it has now expanded to include supercomputers, mainframe computers,
In Section 3, we will talk about the salient features of Linux which makes it a good option for use in servers.
Section 4 talks about the different applications and uses of

Linux operating system.
Section 5 talks about the reasons why Microsoft uses Linux in its servers and the drawbacks of Windows operating system.
Prakruthi c Prakash modify and use for commercial and non-commercial uses. It was initially designed for personal computers, but over the time, it has now expanded to include supercomputers, mainframe computers, servers, and other devices, as it does have a lot Linux features. Linux is the safest server to work upon. That is one of the reasons why even big players like Microsoft use

Linux servers.
Azure Cloud Services, the cloud platform of Microsoft uses Linux based operating system for its functionality. Most of the virtual machines in the Asure Cloud Services run on

Linux Servers.
Section 2 talks about Linux server where we will discuss what exactly is a Linux server, its uses, importance of Linux in brief and why organizations prefer Linux over Windows.

In Section 3, we will talk about the salient features of Linux which makes it a good option for use in servers.
Section 4 talks about the different applications and uses of

Linux operating system.

Section 5 talks about the reasons why Microsoft uses Linux in its servers and the drawbacks of Windows operating system.
Section 6 talks about the future of Linux.
In Section 7, we do a comparative study between Linux and Microsoft to understand the differences between the two.

II. LINUX SERVER

To understand why big players like Microsoft use Linux operating system and servers, we have to first understand the features of Linux Operating System.
Most of the businesses and organizations nowadays want a safe and efficient environment to work with and handle all their data and operations. Linux[2] servers enables to perform all their operations by providing a less rigid interface to work with. Linux is an open-source operating system and the code is available to all to modify and improve its functionality.
Linux is a Unix-like operating system based on the Linux kernel and it is used for smartphones, cars, desktops and home appliances. Linux operating system isn't that popular when it comes to desktop computers but it is the first choice when it comes to powering servers, super computers and mainframe computers in major parts of the world. The reason behind this is that Linux is the most reliable operating system. The reason Linux is more reliable is it has a strong focus on process management, the security of the system and uptime. The issues with Linux operating system are very less when compared with windows operating system. Even though Microsoft has made a lot of recent developments in its operating system, Linux still feels more reliable in most cases.
A study says that 90% of the cloud infrastructures in the world are powered by Linux. This means that Linux is playing huge role in today's world and providing perfect cloud-based solutions is nearly impossible without Linux. Cloud is the present and future. Hence, we can surely say that Linux is here to stay and a future without Linux is impossible.

III. LINUX FEAURES

We saw the importance of Linux in Section 1. Now let's see what exactly makes Linux that effective and reliable. Below features[3] of Linux are discussed to show what exactly makes Linux that safe, efficient and reliable.

A. Open Source

The source code of Linux is free for everyone for free without any restriction on its usage. It is developed by a community of developers around the globe to work on its capabilities and thereby ensuring that the Linux operating system keeps evolving on a daily basis.

B. Security

The most important feature or rather the most significant feature for a user is Safety. The main reason why users prefer Linux is because of its safety. Linux security[4] is one of its most important and top features that makes it unique. The security is provided using user authentication like password protection, the data is secured using data encryption and the access to specific files is controlled. The Linux servers provide a high level of security by having an efficient firewall and a very strict file access system.

C. Stability and Durability

Linux servers rarely crash and they are stable. They are reliable, meaning, they can run for years together as long as they are configured properly. They don't require frequent reboots and the updates don't affect any ongoing services. There is a movie scene which depicts a rocket launch workstation where the analysts in the workstation use Microsoft Operating System. When the rocket is launched, the trajectory of the rocket has to be monitored by a system and al of a sudden, the system automatically starts updating itself. The scientists there are very much annoyed as the update would take about forty-five minutes whereas the rocket land in the next fifteen minutes. Though this is just a funny portrayal of the problem, the struggle is real. The organizations will have to function continuously and they can't afford to wait for their systems to complete its update to carry on their tasks.

D. Multi-programming and Multi-tasking

The inbuilt technical strength and operating system design of Linux makes it possible to run multiple software applications at the same time. Linux can also run multiple programs at the same time. Linux servers enable the programs to run in the background while the user works simultaneously works on another program. This feature of Linux makes multitasking in we-servers possible.

E. Economical and Ease of administration

Linux is an opensource operating system and all its versions are available at a much lower cost when compared to other servers. There is no need of additional software for the administrators to set up any additional tasks and this can be controlled and managed remotely across multiple places.

F. Operational Efficiency

High volumes of data can be handled effectively using Linux servers. Linux servers perform really well in networking and connected devices. These are some of the most significant features in Linux Servers. All these together will enable the users to use Linux servers as their go-to solution for operating the servers. Knowing that the servers require a very high level of security, and security being one of the USP's of Linux, more organizations will be maintaining or switching to Linux.

IV. APPLICATIONS OF LINUX

We have seen what makes Linux an effective operating system in Section 3. Now lets discuss about the various applications of Linux[6] Operating System.

A. Lab Equipment

All the PC's controlling sensors and other lab equipment are usually installe with Linux operating system.

B. Embedded Systems

All the robots, telescopes, engines in automobiles and even set-top-boxes are installed with Linux operating system.

C. Smart Televisions

Smart TV's use Linux operating system to run the various applications installed in the smart TV and also to run its various other tasks.

D. Smart Watches

The demand for smart watches has led to the increase of these smart watches and fitness bands these days. These also run on the preinstalled Linux operating system.

E. Routers, Modems and Firewalls

Most routers, modems, and firewalls use variations of Linux that are pre-installed by the hardware company. DD-WRT is the most common Linux open-source firmware for wireless routers and other embedded systems.

F. Security Equipment

Most of the security gadgets like surveillance cameras, electronic locks and digital cameras use Linux operating system.

G. Synthesisers, Audio Mixers and Interface All the audio processing equipment use Linux operating system.

V. PURPOSE OF LINUX IN MICROSOFT

We know that Microsoft by itself is one of the leading developers of an operating system named Windows. When they can make a user-friendly operating system like Windows, why can't they use Windows or make another operating system to run their servers? The answer for this can be obtained by understanding the drawbacks of windows.

A. Stability Issues

The server is not stable as there are downtime issues encountered on a regular basis. The entire functionality of the server comes to a halt when this happens. There is also a chance where the server will reboot all of a sudden and there will be an unwanted downtime.

B. Speed

Windows servers are not as fast as Linux servers and this isn't ideal if the operation requires fast loading time.

C. Additional Storage Space

Windows requires additional storage space, more primary space and bandwidth to function efficiently. All this will eventually affect the processing speed of the server.

D. Cost

Windows isn't a free platform. The cost of hosting a Windows server will be high. To add to this, there will be limitation on the applications that can run on the Windows server. In addition to all this, Windows servers don't include the rights for virtualization as a Hyper-V host.

VI. LINUX VS MICROSOFT

Let us now do a comparative study between Linux and Microsoft[8] to get a clear understanding of the differences between the two operating systems.

Windows Linux

There are 4 types of user accounts namely Administrator, Standard, Child and Guest.

There are 3 types of users in Linux namely Regular, Administrative and Service users. Uses the micro-kernel which takes less space but lowers the system running efficiency than Linux.

Uses the monolithic kernel which consumes more running space.

Files are stored indirectories/folders on different drives like C: D: and E: drives.

Files are ordered in a tree structure starting with the root directory, further branched out to various other sub-directories. Less security Enhanced security Highly compatible – Any software can be installed and run in Windows. Compatibility is less. User-friendly. Becoming user-friendly over the years. GUI-based operating system. Command-based operating

system. Privacy is less and the operations are monitored.

Privacy is high and the operating system is completely yours. Commercial operating system. Open-source. High risk of crashes, auto- updates and down-time. Highly reliable.

VII. CONCLUSION

Here comes the deciding factor, why Microsoft is using Linux[5] servers. The users are generally confused with which Operating System to consider. They do clear research about all the features and factors they would use in their organization. First and the foremost reason is cost, Microsoft is expensive while Linux is open source and available at a lower cost. Microsoft usually charges for all the services and Linux doesn't. Linux is the most secure server that you can find. When it comes to security, Linux overpowers Microsoft. Windows has a very complicated license model that is also a major disadvantage. Taking all this into consideration, Windows decided that Linux is good enough to run their servers as it serves all theirneeds.

VIII. ACKNOWLEDGMENT

I would like to express my gratitude to my reviewer, Dr. Subhashini for her support. I would also like to thank Ms. Surya Pandey for her support and guidance.

IX. REFERENCES

[1]. Barabanov, Michael. "A linux-based real-time operating system." (1997).

[2]. Bauer, Michael D. Building secure servers with Linux. " O'Reilly Media, Inc.", 2002.

[3]. Dietrich, Christian, et al. "Understanding Linux feature distribution." Proceedings of the 2012 workshop on Modularity in Systems Software. 2012.

[4]. Loscocco, Peter, and Stephen D. Smalley. "Meeting critical security objectives with security-enhanced linux." Proceedings of the 2001 Ottawa Linux symposium. 2001.

[5]. Mary Branscombe, "What is Microsoft doing with Linux? Everything you need to know about its plans for Open Source".

[6]. Yodaiken, Victor, and Michael Barabanov. "Real-time linux applications and design." Proceedings of the Linux applications development and deployment conference (USELINUX). 1997.

[7]. Economides, Nicholas, and EvangelosKatsamakas. "Linux vs. Windows: A comparison of application and platform innovation incentives for open source and proprietary software platforms." The Economics of Open Source Software Development. Elsevier, 2006. 207-218.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Challenges and Advantages of Framing Computer Organization Course

Mehaboobsubhani Jamadar[1], Mohammad Huzaifa[1], Shreedhar V Kalahal[1], Mohammed Ayaan[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Teaching Computer Organization and Architecture to the students of Computer Science and Engineering is aresponsible task, as this course lays foundation for various othercourses like System Software, Principles of Compiler design andOperating System. But it's always been a challenge for the courseinstructor to teach the course effectively. The students ofComputer Science and Engineering have been inclined towardssoftware courses and hence teaching a course without associated laboratory, has been a drawback. To aid students with betterunderstanding of concepts, course instructors decided to redesignthe course. The tutorial was planned to address the previouslymentioned challenges. The most popular RISC ARM processorwas chosen as a case study. To measure the effectiveness of thetutorial, activities were conducted and evaluated as per designed rubrics. In this paper we highlight the challenges and the advantages of tutorial. As a result, we have noticed that studentshave shown better understanding of the courses.

Keywords: RISC processor, ARM, LPC2148, Operating Systems, Computer Architecture, RISC-V.

I. INTRODUCTION

With the transformation from college to university the responsibility of an educator has increased from course design to course evaluation. Modern system engineers should know the concepts like, HPC, Quantum Computing and Parallel Computing, which falls under the system engineering vertical of the department. These courses require strong foundation of the hardware related courses like Digital System Design and Computer Organization
and Architecture. Emergence of Internet-of-Things brings a whole new class of applications and higher efficiency for existing services. Application-specific requirements, as well as connectivity and communication ability of devices have introduced new challenges for IoT applications [1].

Currently industry requires engineering graduates having multidisciplinary skills. In order to cater to these requirements, courses and course contents are revised and redesigned. As observed there is an increased demand for IoT and embedded system based products.

In order to well equip computer science students with current industry requirements, the courses, Digital System Design, Computer Organization and Architecture, Advanced Computer Architecture,

Applied Parallel Computing are included in the curriculum [2].

The course Digital System Design is taught at 3rd semester level which contributes prerequisites needed for the course Computer Organization and Architecture [3].

The courses Advanced Computer Architecture and Applied Parallel Computing are taught at 3rd year. The contents of Computer Organization and Architecture was planned to serve as prerequisites for above courses which are taught at higher semesters. This course was earlier taught without hands-on approach which proved insufficient to grasp and less interactive to majority of the students.

The white paper on "Learning in the 21st Century: Teaching Today's Students on Their Terms", by International Educational Advisory Board (IEAB) Discusses about nature of current generation students and role of educators. Today, information and communications technologies (ICTs) infiltrate classrooms around the world at an exceedingly rapid pace.

In the wake of this influx, educators face growing challenges as they teach a very "wired"—and more and more "wireless" generation of students using technology that is evolving every day. Millennial students do not want to be bound by traditional schedules, and they do not necessarily want to sit in a classroom to learn or in an office to work. Instead, they prefer to use technology to study at any time of the day or night, telecommute from anywhere in the world and define "balance" in their own individual ways [4]. Hence it's a great challenge for the course teachers to be well equipped and be updated with evolving technology.

As the mentioned in Computer Science (CS) curricula recommendation report of IEEE/ACM2001, the course needs to be updated in timely manner to reflect the state of art technology and course topics and contents needs to be carefully adjusted to cover more recent technologies and core topics in greater details [5].This paper discusses how the course contents upgraded to adopt to the new technology ARM and in turn how this ARM helped to address challenges faced during traditional teaching methods to teach to millennial students.

To make the course effective tutorial was planned in such a way that contents of tutorial should reinforce the contents and correlate the ARM architecture with GPU architecture. The tutorial contents were systematically organized to cover the concepts like ARM architecture, ARM assembly programming and application design using embedded C. The OS acts as an intermediary between the user and the computer hardware. The OS provides an abstraction by enabling users to concentrate on doing their necessary tasks or software developers to engineer applications with ease.

With the demand for small, portable, and straightforward OS for deploying on embedded computers, there has been a broad category of operating systems developed for each purpose. In contrast, computer architecture research has tried to simplify computer components for easier integration and better performance. The RISC-V is an open Reduced Instruction Set Computer (RISC) Instruction Set Architecture (ISA) which enables internal modification and design. Berkeley University developed it to reduce the cost of software by permitting more reuse with an open ISA.

The RISC-V implementation can be easily extended by designing custom extensions when additional functionality is desired. The study of OS and computer architecture has been highly dependent on current technology. However, these technologies are either proprietary, hardening internal review, or too complex for academic classes of operating systems and computer architecture. RISC-V breaks such barriers

enabling students and lecturers to enrich their class material further.

A variety of kernels have been ported to RISC-V, including the Linux kernel, Amazon's FreeRTOS, Redox OS, and the Zephyr Project. However, these solutions address the application in embedded systems and do not focus on beginners in the study of operating systems.

This work provides the project and implementation of a basic microkernel for academic purposes, by enabling students and lecturers to dive into computer system research and ease their understanding of computer architecture and operating systems development. This works aims to be used as a laboratory course material for both Operating System and Computer Architecture courses by providing a simple UNIX-based RISC-V kernel.

II. BACKGROUND

Operating Systems

Computer software can be classified into two categories: system programs, responsible for managing hardware resources, and application programs, which execute the actual, work that the user requests. The OS's job is to provide application programmers a clean abstract set of resource interfaces of the hardware, leaving the hardware complexity to the OS, and enabling the programmer

to focus on the application they are developing.

Operating Systems has three main objectives:

(1) They aim to offer convenience, making computer hardware more usable.

(2) They must enable computer system resources to be used efficiently, for example, making sure the memory is efficiently allocated.

(3) They must have the ability to involve, meaning that the OS should be engineered in a way that is easy to extend, develop, test and introduce new features.

The kernel is an essential part of the OS, and therefore, it has special rights comparing to other parts of the system. It is loaded into memory on boot, initializes any required services, and waits for an event. An essential principle in kernel design is the separation of policies and mechanisms, determining what will be done and how it will be done, respectively. This separation is important, as policies are likely to change over time. As systems tend to grow over time, they must be engineered carefully to work properly and be easily modified. The kernel can be engineered as a monolithic kernel or as a microkernel.

Monolithic kernels runs all services such as processes, memory management, and interrupt handling in the kernel space, in a layered approach. The bottom layer is the hardware, and the upper layer is the user interface. Typically, a monolithic kernel is implemented as a single process, with all its components sharing the same address space.

Microkernels removes all nonessential components from the kernel and implements them as system and user-level programs. These services are processes which are known as a server. They encapsulate memory management, process managing, Interprocess Communication (IPC), and so on. The microkernel's main functionality is to provide a communication mechanism between the client program and the servers. If the client requires to read a file, it must interact with the file server through communicating indirectly by exchanging messages with the kernel.

Computer Architecture

Although there are different distinctions made between computer architecture and organization, the former refers to what systems and application

programmers see, which are the attributes that have a direct impact on the execution of a program, where the latter refers to the operational unit and its interconnections that make the architectural specifications. The ISA, the numbers of bits used to represent data types, the Input and Output (I/O) structure, and approaches for memory addressing are all organizational issues that need to be structured.

The computer organization creates a hierarchy of hardware attribute details transparent to the programmer, such as the interface between the computer and peripherals, the memory technology used, the type of processor and control signals.

Computer architectures can be classified in Reduced Instruction Set Computer (RISC) and Complex Instruction Set Computer (CISC). RISC architecture provides a large number of general-purpose registers, the use of a compiler to optimize register use, a simple and limited instruction set, and optimizes the instruction pipeline. In a CISC ISA, the compiler needs to do very little work to translate high-level language into assembly level. However, though it generates smaller code sizes, it needs higher cycles per second.

RISC-V

RISC-V is an open RISC ISA offering both 32- and 64-bit support. The RISC-V's ISA is modular, having a base architecture and several available extensions, enabling multiple variants to coexist. The base architecture is the RV32I, which will not be altered to enable programmers to rely on a stable architecture. The modules are standard extensions on which the hardware designers can choose whether to include them or not. The modular architecture enables small-scale applications with little energy consumption to be deployed with the required extensions to run.

RISC-V supports multiple software stacks, from executing a single application running on an Application Execution Environment (AEE) to multiple programmed OSs running on a single hypervisor. RISC-V provides three modes of execution: Machine-mode (M-mode), User-mode (U-mode), and Supervisor-mode (S-mode), having M-mode the only mandatory mode. Execution modes enable programmers to protect different software stack components and guide programmers on building secure systems. An operation not permitted by the privilege level will cause RISC-V to raise an exception, trapping into the underlying execution environment. Code executing in M-mode is commonly trusted, as it has low-level access to the machine implementation. The U-mode and S-mode are indented for conventional applications and OS, respectively.

III. COURSE DESIGN

In the curriculum of KLE Technological University course contents play crucial role, so they need to be carefully chosen, updated and reorganized by rigorously modifying syllabus. As per the observation the curriculum of various universities show that more focus is given to computer organization. Simulation is used as an aid to teach the organization concepts effectively. But this approach does not give strong support to teach Computer Architecture as compared to hands on approach, where the concepts like addressing modes, Instruction Set Architecture and pipelining are involved. Hence to motivate the students and conduct the course effectively hands-on approach was adopted.

The course content was designed with the following objectives. The Course Outcomes (CO) are as follows:

1. Explain architecture and organization of modern computing systems.
2. Analyze the effect of pipelining and the way it can speed up processing.

3. Illustrate the different ways of communicating with I/O devices and standard I/O interfaces.
4. Analyze instruction execution and data processing in terms of clock cycles.
5. Discuss architecture and operation of a dedicated processor.
6. Write assembly language and embedded C programs for the defined task.

The different activities are planned to assess all the COs. Course outcomes were analyzed to measure the effectiveness of the changes brought out in the course design.

The course contents for the course COA are as follows:

1. Basic GPU Structure Operating System and Principles of Compiler
2. Pipelining Design
3. Input /Output Organization
4. The Memory System
5. Arithmetic
6. The ARM architecture System Software, Mini, Minor and
7. The ARM Instruction Set Capstone Projects
8. ARM Assembly Programming

It discusses GPU structure which describes the internal details of general purpose computing. It also discusses significance of pipeline design and hazards handing. It discusses I/O organization where in procedure of accessing input and output devices to processor plays significant role.

It focuses on memory system, covering the hierarchical architecture of the memory system, types of memory like SRAM DRAM, cache, virtual memory and address mapping.chapter5 discusses how arithmetical operations inside a computer are performed. Rest of chapters cover architecture of ARM and Assembly language programming.

A. Selecting ARM as a case study

It is impossible to teach a course on computer architecture without introducing a computer. A specific computer architecture is required as a vehicle to teach about registers, addressing modes, instruction types, and so on for teaching the concepts effectively to Computer Science students [6]. This paper has suggested that the ARM is an ideal choice for both the professor and the student because it is easy to understand initially, teaching material is widely available, and high-quality PC-based ARM simulators are can be freely downloaded from the Internet. The advantage of an architecture that allows students to get going from day one can't be over-estimated [6][7].

The ARM architecture was earlier taught in the course Embedded System for four credits with an associated laboratory, which is removed from the current KLE Tech curriculum. The Embedded system course consisted of ARM Architecture, Assembly programming and Interfacing the peripherals using embedded C programming. Students come to department with the knowledge of interfacing peripherals to AURDINO in the Engineering Exploration course in first year. Hence they do have basic foundation of interfacing and modelling.

As observed it was found that the curriculum lacks an architecture of one processor which is needed to CS students. [8] To bridge the gap, architecture ARM was selected as Case Study for the course. Also care was taken to eliminate redundant topics which was covered in earlier semesters. Tutorial was planned to teach the architecture effectively.

B. Tutorial Plan

The tutorial was planned to teach one specific architecture with addressing modes, ISA, assembly programming and interfacing ARM with peripherals. The aim was to give the students a platform to explore the hardware details and get a proper insight with the

internal functioning of a Computer. The concepts taught in the tutorial targeted to teach the concept like RISC, CISC comparison, Pipelining concept, memory organization, Assembly and Embedded C programming.

Pipelining concepts incorporated as it should give exposure for students to understand concepts like concurrent execution and superscalar execution. These concepts are the basis for understanding the courses like parallel computing. Journal of Engineering Education Transformations, Special Issue, eISSN 2394-1707. Memory organization of ARM are correlated with modern computing machine as this also serves as basis for operating system subjects. The tutorial was planned with the following objectives as shown in Table 2.

Tutorial Plan			Marks: 20	
Objective: • Study the internal components of a 32 bit system. • Compare various architectures with ARM architecture. • Develop an embedded system based application using ARM processor by choosing suitable IO peripherals for a specified task.				
CO: 1,5 & 6				
Evaluation Criteria and mapping with PI				
Evaluation Criteria	Weightage in Marks (Total 20)	Blooms Level	Timeline	Indicator
Identify the internal components of computer.	5	L2	Review 1 Week 6	1.3.1
Compare a specific architecture with ARM architecture.	5	L3	Review 2 Week 8	1.3.1
Identify the problem for a given theme and write the clear definition and the objectives.	3	L3	Review 3 Week 7	2.1.1
Design and implement the prototype for the identified problem.	7	L3	Review 4 Week11	2.1.2 5.1.1 9.3.1

IV. METHODOLOGY TO CONDUCT TUTORIAL

The tutorial was 2hrs session every week where concepts ARM RISC Architecture was covered to reinforce the core concepts of COA. The activities were planned and conducted as mentioned above.

A. Demonstration on internal components of a computer

This was team activity where students were given demonstration on internal hardware details [9]. This activity helped students to identify, assemble the different components of a computer and visualize how different units like memory, buses, processor and NIC card are organized in a CPU. This activity was assessed for 5 marks to meet GA1.

B. ARM Architecture

Since power consumption is becoming a major concern for current and future projects it is of great importance to maintain adequate performance, while meeting modern power constraints. [10]. In this paper, author discuss about the discuss about the difference in CISC and RISC processors regarding ISA, performance and power consumption. This activity provides a comparison to be able to cover as many software and hardware applications used by different architectures as possible. After the activity, the students will have enough information on the RISC and CISC processors, the differences between them, and where they are most effectively used. This in turn willhelp in determining which ISA of the selected processor is most suitable for the required application [11].

In this session of tutorial, students were given with Exposure to ARM architecture. Differences between RISC and CISC, Features inherited from RISC by ARM, ARM programming model, different working modes and ARM architecture was taught. To familiarize students with different architecture activity was conducted for 5 marks to meet GA1. It was a team activity where each team of four students were asked to present comparison of ARM architecture different architecture.

C. Assembly instructions.

ARM assembly programming gives positive impact to study the Computer Organization and Architecture [6]. It also give better exposure for students to design assembler in the System Software course taught in

later semesters. This course also serves as a pre requisite for Compiler Design course. These sessions involved teaching the following concepts to help students to write assembly programs:

- Addressing modes
- Instruction classification
- Data transfer
- Data processing
- Control flow
- Usage of tool: Keil IDE

This activity helped students to get familiarized with how instructions get executed, how data is read and written to a memory and interaction of assembly code at hardware level. To make the tutorial successful different activities were planned, conducted and assessed to meet the chosen Graduate Attributes. As this course provides fundamental knowledge to many other courses GA1 was chosen. To assess the effectiveness of tutorial simple application design was used, which meets the graduate attributes
GA2, GA5 & GA9.

D. ARM interfacing.

These sessions were planned to expose students as to how ARM can interface with different peripherals like LED, LCD, motors and hex-keypad[11]. This activity helped students to build simple applications.

E. Develop an application using ARM processor.

In this activity students were asked to form team which consisted of four members. Different themes were planned and given to all eight batches

i. Smart city
ii. Agriculture
iii. Renewable energy
iv. Aid for old age and physically handicapped
v. Transportation
vi. Aid for education
vii. Security
viii. Automobiles

This activity helped students to design a computer based system to monitor and control the peripherals. This activity addressed GA2, GA5 and GA9. It was assessed for 10 marks.

F. Rubrics for various planned activities

Rubrics for assessment of various activities were planned as shown in tables 4 and 5.

Table 4: ARM Architecture Comparison Rubrics

Criteria	Excellent(10-8M)	Good (7-5M)	Average (4-1M)
Literature survey	• Very thorough details of ARM architecture and given processor architecture were collected. • Performance parameters of ARM in comparison with other processors are clearly stated ✓ Power consumption ✓ Processor speed ✓ Pipeline stages ✓ Instruction set ✓ Memory hierarchy ✓ Word size of a processor.	• Details of ARM architecture and given processor architecture were collected. • Performance parameters of ARM in comparison with other processors are stated. ✓ Power consumption ✓ Processor speed ✓ Pipeline stages ✓ Instruction set ✓ Memory hierarchy ✓ Word size of a processor	• Details of ARM architecture were collected but given processor architecture details were not complete • Performance parameters were not clearly stated.
Presentation	• Presented the contents with confidence and without much referring to slides. • Eye contact with audience was maintained. • All queries put forth by the audience was handled gracefully.	• Presented the contents without much referring to slides. • Eye contact with audience was maintained occasionally. • All queries put forth by the audience was handled.	•Presented the contents without confidence and oftenly referred slides. •Most of queries put forth by the audience were not answered.

Table 5: Model of ARM Application Rubrics

Criteria	Excellent(10-8M)	Good (7-5M)	Average (4-1M)
Presentation on Problem identification	• Effectively identified problem. • Identified desired solutions or options • Topic introduced clearly and in an interesting way. Purpose of talk was made clear. Outline of points was given • Presenter spoke clearly and at a good pace to ensure audience comprehension. • Delivery was fluent and expressive.	• Problem Statement identified but no clarity about solution • Topic introduced clearly, and purpose of talk was made clear. • Presenter usually spoke clearly to ensure audience comprehension. • Delivery was fluent.	• Uncertainty in problem identification. • Problems not stated clearly • Topic introduced and the purpose of talk was not very clear. •
Model Implementation and Demonstration	• Great care taken in construction model so that the structure is neat, attractive and follows plans accurately. • Significant gain in knowledge or skills. • Model works fine for all possible inputs. • States conclusion with justification based on reasonable data interpretation of data	• Model accurately followed the plans, few details could have been refined for a more attractive product • Partial gain in knowledge or skills. • Partially working model • States conclusion based on reasonable data interpretation of data	• Model appears careless or haphazard. Many details need refinement for a strong or attractive product. • Model not working • States no conclusion or with conclusion based on reasonable data interpretation of data

The weightage in marks to all the CO's are distributed as per the relevance of topics and how they map to further courses in higher semesters. The weightage in marks planned for the CO's to be addressed in our course is as shown in Fig 1.

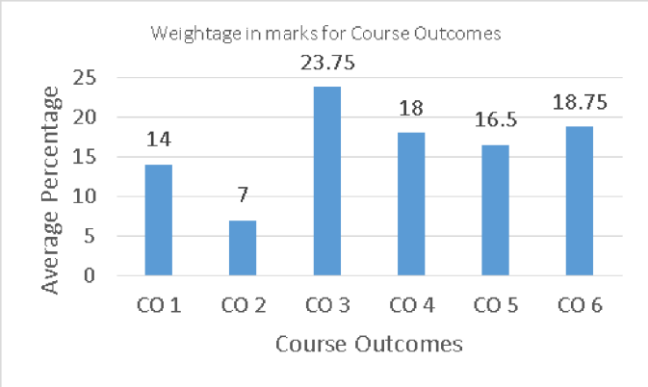


Fig. 1. Marks weighatge for Course Outcomes.

G. Samples of ARM Application developed

Some of most interesting models designed using ARM were noted and are as follows Automatic Seed Sowing Machine: To design a system that helps in sowing the seeds at fixed distance Objectives:

- ✓ The Device should be automatic.
- ✓ It should sow the seeds in the required pattern.
- ✓ Amount of seeds to be sowed can be controlled by inputs

Fig 2 shows working model for seed sowing machine. This machine developed by students was demonstrated as to show how automatically seed sowing can be done to aid the farmers.

Fig. 2. Demonstration of seed sowing machine

Automatic Crop Cutting Machine: Automatic crop cutting system with IR sensor.

Objectives:

- ✓ The Device should be automatic.
- ✓ Able to detect crops at specified height and cut the crops.
- ✓ Able to detect and stop the machine if object is not crop.

Fig. 3. Model of crop cutting machine

Fig 3 shows working model for crop cutting machine. The machine is designed such way that it should cut only plants and if any other obstacles found it should stop the machine.

V. RESULTS AND ANALYSIS

This work was proposed first time in KLE University, hence results cannot be compared with previous batch. But the planned Cos attainment and POs attainment for the course is as shown in the following diagrams. The attainment of results for all the planned activities for Internal Semester Assessment (ISA) and End Semester Assessment (ESA) is explained below.

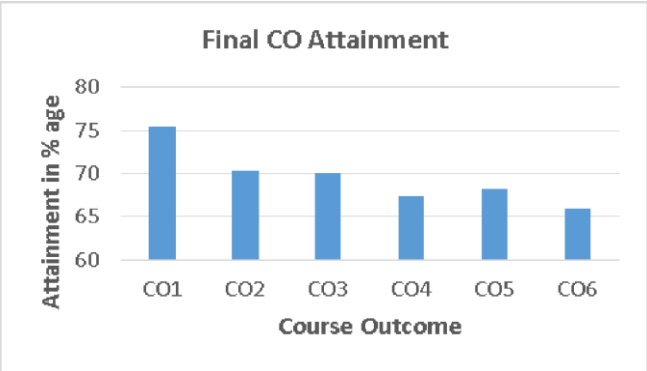


Fig 4: Final CO attainment

VI. CONCLUSION

The study of COA focuses on the interface between hardware and software, and emphasizes the behavior and structure of the system. There is a fundamental relationship between hardware, programming and software components in computer systems. In order to be an efficient programmer, it is very important to understand the computer system as a whole. Tutorial on the COA was very effective to teach the theoretical course which was otherwise considered to be a dry course. All the activities were carried out successfully in the tutorial according to the rubrics designed. Inspite of hectic schedule students enjoyed doing the activities. The concepts taught and the activities conducted serve as a prerequisite for other courses like Advanced Computer Architecture, System Software and Parallel Computing. These activities can help the students who are seeking to carry out the projects in IoT and other areas of Embedded System Applications.

In order to motivate students to take up projects in IoT, the authors strongly believe that advanced ARM processor (Raspberry Pi) can be included in the tutorial.

VII. REFERENCES

[1]. Samie, Farzad, Lars Bauer, and Jörg Henkel. "IoT technologies for embedded computing: A survey." In Hardware/Software Codesign and System Synthesis (CODES+ ISSS), 2016 International Conference on, pp. 1-10. IEEE, 2016.

[2]. Bourque, Pierre, and Richard E. Fairley. Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0. IEEE Computer Society Press, 2014.

[3]. Calazans, Ney LaertVilar, and Fernando GehmMoraes. "Integrating the teaching of computer organization and architecture with digital hardware design early in undergraduate courses." IEEE Transactions on Education 44, no. 2 (2001): 109- 119.

[4]. “Learning in the 21st Century: Teaching Today’s Students on Their Terms”,White paper by INTERNATIONAL EDUCATION ADVISORY BOARD.

[5]. The Joint Task Force on Computing Curricula IEEE Computer Society and ACM, Computing Curricula 2001 Computer Science- Final Report, 2001.

[6]. Zhu, Weiying. "Teaching Assembly Programming for ARM-based Microcontrollers in a Professional Development Kit.", IEEE, 2017

[7]. Clements, Alan. "ARMs for the poor: Selecting a processor for teaching computer architecture." In Frontiers in Education Conference (FIE), 2010 IEEE, pp. T3E-1. IEEE, 2010.

[8]. Clements, Alan. "The undergraduate curriculum in computer architecture." IEEE Micro 20, no. 3 (2000): 13-21.

[9]. Jian, Kuodi. "Meet the Challenge of Teaching Computer Organization and Architecture---- Physical Computing."

[10]. Hamacher, C. (n.d.). COMPUTER ORGANIZATION AND EMBEDDED SYSTEMS (6th ed.). McGraw Hill.
[11]. Chevtchenko, S. F., and R. F. Vale. "A Comparison of RISC and CISC Architectures." resource 2: 4.
[12]. El Kady, Sarah, Mai Khater, and MerihanAlhafnawi. "MIPS, ARM and SPARC-an architecture comparison." In Proceedings of the World Congress on Engineering, vol. 1. 2014.
[13]. Furber, S. (n.d.). ARM System-on-Chip Architecture (2nd ed.). Pearson.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Product Recommendations: From Pixels, Cookies, Machine Learning to Advertising

Sonitha Mandava*, S R Meghana Sarvepally, Megana M
*Department of Computer Science and Engineering New Horizon College of Engineering, VTU, Bengaluru, Karnataka, India

ABSTRACT

In today's day and age, businesses are making the best use of paid and targeted advertising to maximize the number of sales in their business by recommending their products to consumers who are most likely to buy them and not show their products to those consumers who are less likely to click on the ads. In addition to this, they also make use of retargeting advertising to show advertisements to consumers who viewed their products. This paper discusses how targeted advertising is done using pixels, cookies, and trackers to help businesses sell their products and how machine learning is used to recommend products to consumers by understanding user behavior.

Keywords: Targeted advertising, Cookies, Pixels, Recommendations

I. INTRODUCTION

An average person takes anywhere between 7 - 21 times of exposure to a product before they buy it. In order for businesses to make the maximum profit they need to advertise their products to potential consumers and recommend new products which they are most likely to be interested in.

This paper examines how these advertising systems work in web. When a person visits any website based on products and later visits some other websites notices that the advertisements related to the products of the previous website appear on the page.

The technology behind this strategy is discussed in this paper which is based on cookies and pixels. Also the machine learning algorithms used in the individual application for generating recommendations plays a crucial role in this and thus is the second aim of this paper.

II. COOKIES AND PIXELS

2.1) COOKIES

Cookies are text files that store the login information about a user after visiting a particular website. A unique id is given with respect to the user and their computer to the data stored in the cookie.

The cookie that is used in the process of recommendations is the HTTP cookie or the internet cookie. These cookies are built specifically for Internet web browsers to track, personalize, and save information about each user's session. A session is a time you spend on a site.

This cookie is shared with the web server which stores the information about the website. This web server later sends the information about the website to your web browser. These cookies are sent to the web browser only if the server wants to send them to the user and are stored locally in the user's computer in the form of key-

2.2) PIXELS

A snippet of tracking code that helps track a user's activity and what type of ads they click is called a pixel. This pixel becomes essential for businesses to help them run targeted ads to consumers who are likely to buy their products.

There are many types of pixels that play a role in the process of tracking the behavior of a user. The most important ones are:

1. A universal pixel - The pixel that tracks the user's behavior on multiple pages on the same website.
2. Retargeting pixels - These pixels are focused on delivering ads to you based on the user's recent browsing history. For example, if a user browses shoes on a website, when he visits other websites, he will receive ads related to shoes in the advertising banner section.
3. Conversion pixels - These pixels track the purchases that happen on the website.

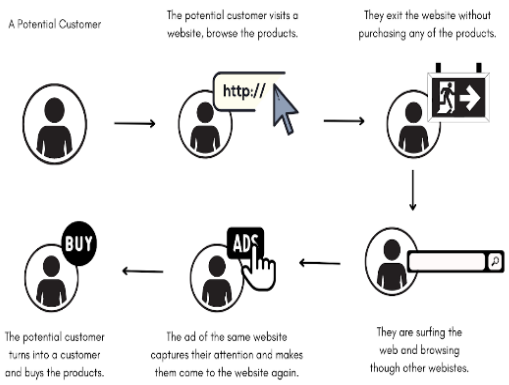


Fig. 1 Retargeting pixels

2.3) COOKIES VS PIXELS

For a business to run targeted advertisements, a pixel needs to be installed on their website. This pixel is provided to the user by advertisement platforms such as Facebook and Google [1].

Let's say a website ABC is created and a user visits it. When the user visits this website ABC, a basic web tracking cookie is placed on their computer. But, this cookie does not communicate with the ad platform created by the website ABC's owner. Thus, for the advertisement system to work the pixel on the website ABC comes into the picture. The advertisement platform's server which has provided the pixel to the website ABC sets a cookie on the user's computer to keep a track of the user.

The advertisement server has access to spaces in other websites. When ABC's owner buys this advertisement space, the specific advertisements related to ABC's products are displayed(retargeted) in these spaces to the targeted potential user-consumer. This is possible only because of the cookie that is placed on the users website through the pixel placed on ABC's website.

2.4) COOKIES LOCALITY

Cookies are files created by the sites visited by the user. Cookies are generally local to a device and do not travel across devices. The advertisements displayed to the user are available only on that specific device and the cookies are based on the sites visited on that device [2].

The advertisements and recommendations that a user gets based on our electronic mail logins are because of the mail cookies. For example, Gmail cookies are present which get created when a user visits a website with their specific Gmail id. Pixels present in those websites track the cookies and all the sites the user visits and display the recommendations in the advertisement space.

2.5) RECOMMENDTIONS THROUGH COOKIES AND PIXELS

When the pixel shares the cookie information to the web server that hold the data of the websites, these web servers analyze the user information and send it to the machine learning algorithms present in the code which in turn decide the data, recommendations to be displayed to the user.

III. PRODUCT RECOMMENDATION USING COOKIES AND PIXELS

While the user behavior is tracked using Pixels, Cookies, and other trackers such as Google Tags, in order for the user to see advertisements they would potentially be interested in, a machine learning algorithm becomes vital for this.

The machine learning algorithm's role is to analyze user behavior from the data given and what type of users are interacting with a certain advertisement.

When the tracker such as Facebook Pixel and Google Tag tracks user behavior, this data is given as an input to their respective machine learning algorithm which analyzes the data and shows the users products they have viewed or they would potentially be interested in.

Let us discuss a few of the machine learning algorithms used for product recommendation.

3.1) CONTENT-BASED FILTERING

In Content-Based Filtering, the machine learning algorithm takes into account the product features and compares user interests to the features. If the product has the most overlapping features with the user's interests, then this product is recommended to the user.

Thus, the features are extracted from the user profiles based on the items/content they have consumed in the past [4]. Content Based Filtering is user specific. Only the products having positive reviews are recommended to the user. The product description is used by the system to learn about the kind/type of products the user is willing to buy. All the items to be recommended to the user are stored in a database and a unique id is provided to each item [5]. The efficiency of this algorithm is based on the description available for the products.

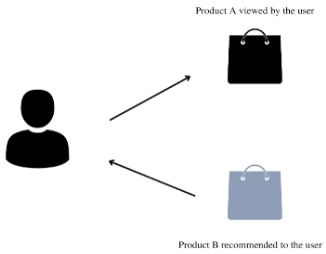


Fig. 2 Content Based Filtering

Various machine learning techniques could be used to implement this. Td-idf (term frequency- inverse document frequency) weights can be used for finding the similarity between documents. We can use machine learning algorithms for text categorization [6]. Naïve bayes classifier, Decision trees, fuzzy systems, neural networks can be used.

Content based filtering is most successful in recommending web pages, news or publications. But it is difficult and doesn't make sense when we need to compare a large number of records. Music, movie recommendations are impractical. The scope of recommendation, new recommendations is restricted as it is based on similarity.

3.2) COLLABORATIVE FILTERING

In collaborative filtering, the interests or preferences of multiple users are taken into account and automatic predictions are made. "*These Systems aggregate data about customers purchasing habits or preferences and make recommendations to other users based on similarity in overall patterns*" [7]. The predictions here are also user-specific but they also use the data collected from other users as well.

For example, In case Person A and B have both purchased the same product, a different product used by either Person A or B is recommended to the other person because they are similar users.

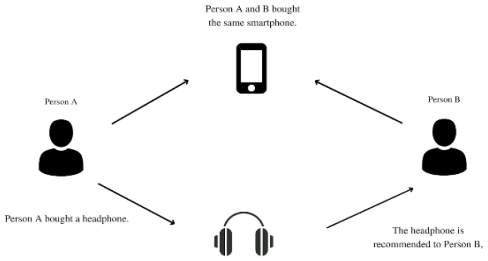


Fig. 3 Collaborative Filtering

Thus, in collaborative filtering we recommend users the items which are liked by the users of similar interests.

To find out the users possessing similar interest we can use cosine distance between two users in the utility matrix. As in a database we can have null values, while using the above method we can give zero value to all of those unfilled columns. If the obtained cosine is larger implies that the angle between the users is small and thus the similar things can be recommended.

		Product1	Product2	Product3
User1		6	5	-
User2		2	4	-
User3		-	2	1

The M-distance [8] is defined as the difference between the average ratings of two products and is represented by mi. The missing ratings are not considered while calculating the average. The time complexity for computing the average rating of all items is O(mn). The M-distance between products mi and mj is given as $mdi,j = | \bar{}ri - \bar{}rj |$.

The other methods which can be adopted are normalizing ratings where we take count on the ratings, adding if it is positive or subtracting if negative and thus finally we end up with either with positive or negative values [9]. Using those values we make the clusters of users.

In the popularity-based recommendation system, the most popular or trending products are chosen and recommended to users.

If a product is bought by many people then it is considered to be popular and is recommended to other users.

It is a memory based system in which the prediction is based on the similarities of the matrix which is computed previously [13]. Viewing events and purchasing events is feeded to the training algorithm and a model is deployed to the search engines which recommend the recommendations present in the database to the users.

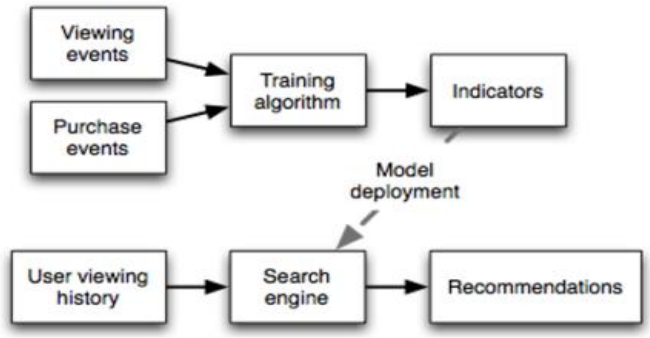


Fig. 4 Popularity based system logic

IV. ADVERTISING USING PRODUCT RECOMMENDATIONS

After tracking the users' behavior through cookies and pixels and sending the data to the machine learning algorithm, the data of product recommendations are used for advertising the products to the user.

Any pixel that is being used needs to be connected to a specific advertising platform. Here are the most popular pixels, and which platforms they advertise on.

PIXEL	PLATFORM
Facebook Pixel	Facebook and Instagram
Google Tag	YouTube and Websites on Google
Bing UET Pixel	Bing

Table 3 Pixel-Platform

Based on what pixels are installed on the websites and what advertisement platform they are connected to, the ads are displayed in the respective places.

For instance, if only a Facebook Pixel is used on a website, the advertisements can only be made on Facebook and Instagram. On the other hand, if a Google Tag is the only pixel used, the advertisements can't be put on Facebook or Instagram - only YouTube and Websites on Google are allowed.

4.1) ADVERTISING ON WEBSITES

Website owners can choose whether or not they want to display ads on their websites. In case they do, there are locations they can choose to display ads - on the top of the website, on either side or on the bottom. A dynamic advertisement space is created at the location of choice where the advertisements will be displayed based on the users' interests. This can be done through Google Ads Manager and by installing a Google Tag.

4.2) ADVERTISING ON YOUTUBE

YouTube has multiple locations they can place ads and they allow multiple forms such as both video and static advertisements. Creators or YouTube can choose where and what ads will be displayed. An example is a food delivery advertisement that is shown before the beginning of a video. This can be done through Google Ads Manager.

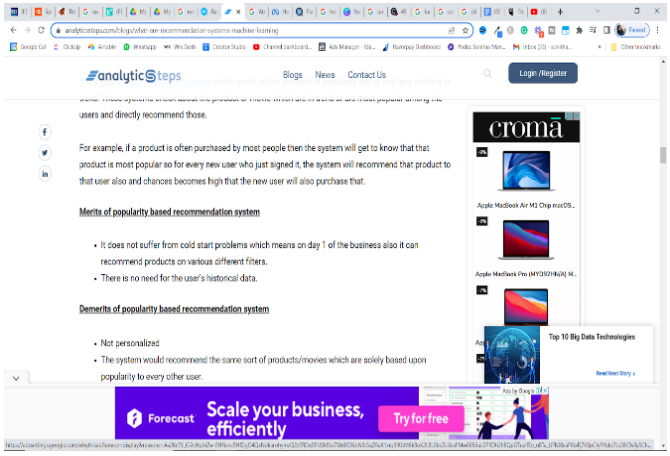


Fig. 5 Advertisements on web

4.3) ADVERTISING ON INSTAGRAM AND FACEBOOK

In order for businesses to advertise on Instagram and Facebook, they need to use a Facebook pixel and use the Facebook ads manager. Businesses can choose where and in what form they want to display ads - Facebook stories only, Instagram stories and feed, and so on.

V. CONCLUSION

Product recommendation is one of the best ways to help businesses get more sales by making use of machine learning along with cookies and pixels, businesses are making use of the potential of advertising,

By making use of trackers such as cookies and Facebook Pixel, businesses are able to track users and understand their behavior. This data is analyzed by machine learning algorithms such as Content-Based Filtering and Collaborative filtering and product recommendations are being suggested. Based on the product recommendations, a relevant advertisement is shown to the user making it very likely that they would click on it.

VI.REFERENCES

[1]. https://learnwebanalytics.com/whats-the-difference-between-a-cookie-a-pixel-and-a-tag/

[2]. https://www.kaspersky.com/resource-center/definitions/cookies

[3]. Jonathan R. Mayer and John C. Mitchell. 2012 Under license to IEEE. DOI 10.1109/SP.2012.47. Third-Party Web Tracking: Policy and Technology

[4]. F.O. Isinkaye, Y.O. Folajimi and B.A. Ojokoh. "Recommendation systems: Principles, methods and evaluation" published and made available on 20 August 2015.

[5]. Michael J. Pazzani and Daniel Billsus. "Content-Based Recommendation Systems" in May 2013.

[6]. Raymond J. Mooney and Loriene Roy. "Content-Based Book Recommending Using Learning for Text Categorization."

[7]. Robin Burke. "Integrating Knowledge-based and collaborative-filtering Recommender Systems" in the proceedings of the Workshop on AI and Electronics 1999.

[8]. Mei Zheng, Fan Min, Heng-Ru Zhang and Wen-bin chen. "Fast recommendations with the M-Distance" published in 2016.

[9]. Geeks for geeks organization presenting with information on various fields.

[10]. Basilico, J.; Hofmann, T. Unifying Collaborative and Content-Based Filtering. In Proceedings of the Twenty-First International Conference on Machine Learning—ICML '04, Banff, AB, Canada, 4–8 July 2004; p. 9.

[11]. Hyeyoung Ko, Suyeon Lee, Yoonseo Park and Anna Choi. "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Appication Fields" Electronics 2022, 11, 141. https:// doi.org/10.3390/electronics11010141.

[12]. S. G. K. Patro et al., "A Hybrid Action-Related K-Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems," in IEEE Access, vol. 8, pp. 90978-90991, 2020, doi: 10.1109/ACCESS.2020.2994056.

[13]. KeshettiSreekala. "Popularity Based Recommendation System" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Detailed Analysis of The Metaverse's Opportunities and Difficulties

Vamshika Sushma Appaji*, Vemireddy Reddy Dheeraj Reddy, Duddukunta Maneesh Reddy, S. Yerrinatha Reddy

*Department of CSE, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Since the popularization of the net within the Nineteen Nineties, the Internet has unbroken evolving. we've created various computer-mediated virtual environments as well as social net-works, video conferencing, virtual 3D worlds (e.g., VR Chat),augmented reality applications (e.g., Pokémon Go), and Non-Fungible Token Games (e.g., Upland). Such virtual environments, albeit non-perpetual and unconnected, have bought U.S.A. various degrees of digital transformation. The term 'metaverse' has been coined to more facilitate the digital transformation in every side of our physical lives. At the core of the metaverse stands the vision of Associate in Nursing immersive net as a big, unified, persistent, and shared realm. whereas the metaverse could seem futuristic, catalyzed by rising technologies like Extended Reality, 5G, and Artificial Intelligence, the digital 'big bang' of our Internet isn't far. This survey paper presents the first effort to supply a comprehensive framework that examines the most recent metaverse development below the size of progressive technologies and metaverse ecosystems, and illustrates the chance of the digital 'big bang'. First, technologies AR the enablers that drive the transition from this net to the metaverse. we tend to thus examine eight enabling technologies strictly - Extended Reality, User Interactivity (Human-Computer Interaction), Artificial Intelligence, Blockchain, laptop Vision, IoT and artificial intelligence, Edge and Cloud computing, and Future Mobile Networks. In terms of applications, the metaverse system permits human users to measure and play among a self-sufficing, persistent, and shared realm. Therefore, we tend to discuss six user-centric factors –Avatar, Content Creation, Virtual Economy, Social acceptableness, Security and Privacy, and Trust and responsibility. Finally, we propose a concrete analysis agenda for the event of the metaverse.

Keywords—Metaverse, Immersive Internet, Augmented/Virtual Reality, Artificial Intelligence, Digital Twins, Networking and Edge Computing, Virtual Economy.

I. INTRODUCTION

WHAT PRECISELY IS THE METAVERSE?

The metaverse might be a massively scalable, persistent network of interconnected virtual worlds where people can work, socialize, transact, play, and even generate in real time. It employs cutting-edge virtualization and technology (AR, VR, perception sensors, and so on) to completely immerse the user in the virtual environment. This indicates that the user

will travel within an always-present universe, which he will have access to whenever he wishes. Many proponents believe that in the ideal artistic movement version of "The Metaverse," there would be a single platform where you could connect your persona, identity, and platform services, and beneath that, numerous worlds would be constructed to which you could obtain access. You'll be a part of, leave, or maybe join a planet with multiple sub-worlds. The definition of a digital identity, digital possession, digital currencies, and universal interchangeability of digital assets are still crucial aspects in establishing a fully functional economy in a very virtual society. In this sense, the metaverse has the potential to replace many parts of how business is conducted, such as what it means to attend a concert, how to discover art exhibitions, and, most importantly, how people learn, study, move, and even make friends.

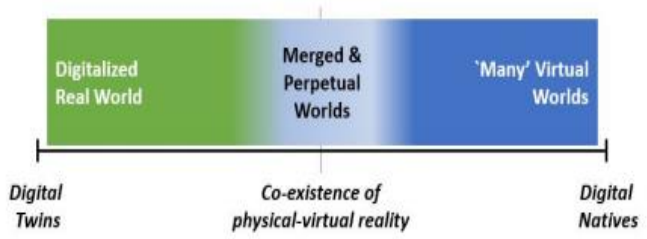


Figure 1: Digital twins-native continuum

II. THE METAVERSE'S EVOLUTION

There are opportunities in practically every market area when you rely on the metaverse's economic science, or metanomics. Consider this: if you have an online avatar and want to change what it/you are wearing, you could be able to purchase limited-edition, digitally branded clothing that you choose while perusing a virtual salesroom. Alternatively, you might start your own small company, such as a virtual personal club or an art gallery where you display your most recent and finest works.

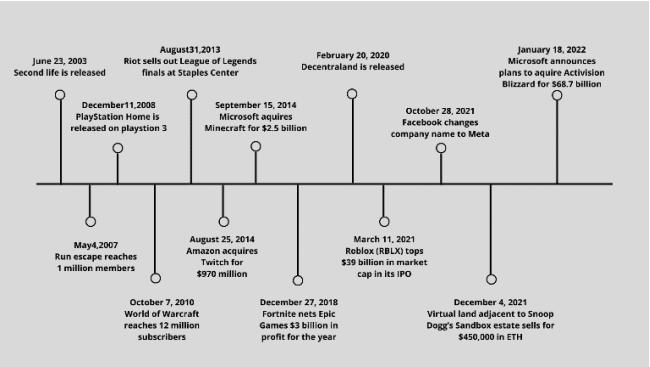


Figure 2: Evolution of Metaverse

III. WHO CONTROLS THE METAVERSE?

Metaverse is a complicated phenomenon that necessitates the integration of complex technology and infrastructure in order to create a 3D virtual environment. The Metaverse is made up of seven verticals that must work together to get the greatest results. What is necessary to make the Metaverse a reality is as follows:

- Infrastructure: High-grade connection, which necessitates WIFI, 5G, and so on.
- Human Interface: Virtual reality and augmented reality gears to join the Metaverse.
- Decentralization: Artificial intelligence, Blockchain technology, and crucial decentralization technologies.
- 3D visual modelling framework created by Spatial Computing.
- The allocation and administration of digital assets as well as e-commerce in the Metaverse are referred to as the Creator Economy.
- Development of a content engine and interaction mode, as well as social media integration.
- Experiments: Adapting previous experiences to new tasks and situations.

All of these layers are necessary for the formation of the Metaverse, which gives the world an immersive experience. Only a few businesses have enough assets and technology to enter the race to build one today.

Epic Games, Facebook (Meta), Niantic, Nvidia, Microsoft, Decentraland, and Apple are among the corporations investing in the Metaverse.

IV. THE METAVERSE FOR BUSINESS

The term 'Metaverse for Business' refers to the construction and implementation of 3D environments for businesses, whether for training, achievement, social events, or connection with shoppers and consumers. There are a few things that organizations will change and manage to meet their needs in the business metaverse: They'll need the authority to demand administration of their own metaverse, even though they'll have their own website. Employees are also inclined to accept the perks on offer from the metaverse, according to research. According to Lenovo, 400 employees are willing to participate in the metaverse and believe it will provide significant advantages. The influence of the COVID-19 epidemic on world civilization will be blamed for causing a shift in worker attitude to the metaverse. The epidemic has thrown a wrench into the traditional 9-to-5 workday. Even once the epidemic is over, it looks that the majority of businesses will be able to accept flexible operating patterns and the ability to make decisions from afar. Clearly, the metaverse has a critical role to play in ensuring that employees remain engaged while also helping companies overcome productivity issues associated with remote working, such as video conferencing tiredness. With the rising use of Augmented Reality (AR) and Virtual Reality (VR) devices, interactive online ecosystems for specialized coaching courses will be developed, delivering virtual tasks and examinations that participants will want to move to in order to pass such coaching. Significantly, AI avatar bots, such as virtual customers or virtual patients, may be employed as a component of Associate in Nursing engaging and exciting learning experience in any industry. Different 3D ecosystems will be built to cater precisely to the needs of businesses, whether through coaching and growth or the staging of online events. Businesses will host and participate in conferences, create goal break places, organize team development sessions, attend company partnership calls, and even produce and advertise events using these 3D event platforms. The metaverse will host or link to any website, picture, video, or PDF, as well as include regions where AR related webpages on the internet are connected to each other. The avatar in the metaverse is a private identity in a world where contact with others is possible, and the engineering science of the setting mirrors the workplace or location where the event is taking place or may appear at any time. In a nutshell, the metaverse will collaborate with or modify reality and ordinary business operations.

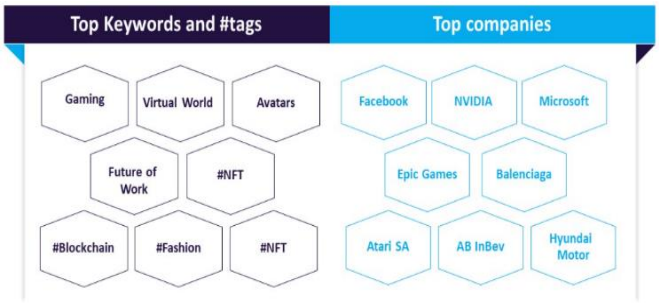


Figure 3: Metaverse Companies

A. Sandbox

Regardless, Sandbox is maybe the most well-known metaverse project. It's an Ethereum-based virtual environment that relies on blockchain technology. Its virtual world makes use of NFTs and cryptocurrencies, as well as its own digital money, the SAND token. The game is similar to a digital sandbox, in which players can host events of any size after registering. Users may create and mint assets, as well as purchase, sell, and trade them. The trade of NFTs is built on the Ethereum blockchain and is dependent on the total

number of tokens in a player's ownership. Users should be forced to create an online persona or Associate in Nursing avatar, which is a Metaverse feature.

B. Metahero

Its idea ensures something unique, such as offering consumers 3D scanning and sculpting services that allow them to duplicate any real-world object. Metahero is considering upgrading to 16k ultra-HD scanners, which would allow users to create ultra-high-definition avatars by digitally replicating real-world objects. The HERO cryptocurrency will be used to buy and sell them.

C. Bloometa

Bloometa was created to be the first-ever comprehensive metaverse virtual system. This may make it easier for them to exchange in-game stuff. It's a cross-chain marketplace where players may buy, sell, mint, and auction NFT assets in metaverse games. Bloometa is one of a kind since it's rethinking how we'll exchange virtual goods in the Metaverse. It's based on the Tezos blockchain, and it's expected to grow into a multi-chain platform that includes Ether, Solana, BSC, Avalanche, and EVM chains.

D. Bloktopia

Bloktopia is one of the most eagerly awaited metaverse releases, with a beta version slated to open in March 2022. It's a skyscraper-themed metaverse project that allows users to arrange grand events and participate in activities within the vertical 21-story virtual structure. Bloktopia has world-famous entertainment venues such as the World Wrestling Entertainment (WWE) and the National Basketball Association (NBA) (NBA). The platform also intends to incorporate NFTs in order to encourage users to build communities and remain devoted participants.

E. Decentraland

Decentraland is a well-known metaverse project whose goal is to create social situations. It's a virtual universe in three dimensions that incorporates the physical world of humans. The MANA token is used. Users in Decentraland can use the MANA cryptocurrency to purchase virtual pieces of land (real estate) as NFTs. Decentraland is currently based on the Ethereum blockchain, with plans to expand its real estate roots to include e-commerce, marketing, and retail in the future. What makes it unique is that the users completely control and govern these sections.

V. HOW THE METAVERSE IS HELPING THE E-LEARNING INDUSTRY

Physical learning has always dominated the educational environment for college students and academics throughout the world. After the pandemic, digital learning took off and spread throughout the country, transforming the traditional learning paradigm into one that is heavily reliant on technology. The metaverse, on the other hand, has shown to be beneficial to businesses and marketers. It is undeniably true that the metaverse will benefit the eLearning industry.

A. Better eLearning And Playing Environment

In today's world, a student prefers to study more with his smartphone than with a book. Metaverse applications will provide a virtual home for college students where they may roam about, take notes, and chat with other students, completely revolutionizing the eLearning industry. They will play games in a virtual environment that closely mimics reality at the same time. Additionally, students may use the projected applications to change their clothes, haircuts, and emotions, among other things.

B. Teachers' Better Illustrations for Students

Although many teachers utilize video-calling applications, they are unable to provide an accurate representation of real-life items using them. Instructors will be able to successfully convey such pictures to their pupils' using technologies like AR. For example, if a teacher wants to show the components of a car, they may utilize holographic software to reflect a 3D picture. Scientific and mathematical experiments can be better understood by students.

C. 3D Visualizations for Better Learning Resources

With 3D representations, the metaverse provides more tools for understanding. Students will be able to go deep into the books, hear the text, and visualize diagrams in a very 3D manner, thanks to the use of virtual reality. For historical themes, the VR will display academics animated movies for better study. Exams will be far more interactive if they are integrated with metaverse apps. For example, test questions will be interactive, and students will be given specific case studies that are virtual yet seem genuine. In this way, the metaverse will create learning resources, transfer them to reality, and create a stronger eLearning market.

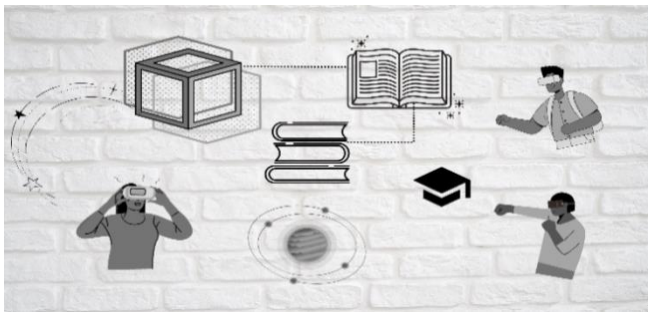


Figure 4: E-Learning in Metaverse

eLearning apps AR taking the mode of learning from offline to on-line models. With the incorporation of VR and AR technologies into these apps, we are able to move one step nearer to the look of the metaverse. it'll take nearly a decade to remodel the eLearning trade into a metaverse. Metaverse is coming and World Is not Ready for it The metaverse is resurfacing. It was originally a science-fiction vision, most famously in Neal Stephenson's novel "Snow Crash," of an all-encompassing virtual universe that might live aboard the real one. However, technological advancements have brought this transition of human civilization close enough to reality for us to consider its implications. In the metaverse, a user could worship a digital avatar, similar to a character in a video game. They may experience a digital universe as dynamic and intriguing as the actual one via the eyes of their avatar. Some futurists believe that we would now go to doctor's appointments or be classified there the metaverse is resurfacing. It was originally a science-fiction vision, most famously in Neal Stephenson's novel "Snow Crash," of an all-encompassing virtual universe that might live aboard the real one. However, technological advancements have brought this transition of human civilization close enough to reality for us to consider its implications. In the metaverse, a user could worship a digital avatar, similar to a character in a video game. They may experience a digital universe as dynamic and intriguing as the actual one via the eyes of their avatar. Some futurists believe that we would now go to doctor's appointments or be classified there.

Today, there are glimpses of the metaverse everywhere. Virtual concerts draw record crowds, high-end designers sell virtual clothing, and leisure has become a source of nourishment for people all over the world. Immersive games like Fortnite, Minecraft, and Roblox, where players may communicate, search, and attend events in a highly simulated environment, are some of the closest corollaries to a full-fledged metaverse. On-line multiplayer games have previously been shown to affect the spread of disinformation and conspiracy theories. In-game communication capabilities will be

used by players to spread rumors or "false news," targeting others in difficult-to-trace ways.

VI. CONCLUSION

We may expect to see the plug around the metaverse settle in the coming years, bringing true breakthroughs back to the fore. The metaverse provides the answer to the current dilemma, particularly in terms of operational habits, and hence the necessity for brand new ways to incorporate virtual worlds into prior routines. The benefits and prospects of the metaverse, and hence the metaverse for business, are limitless: and alternative possible barriers like accessibility and variety are less of an issue than they look for both large and small businesses. If companies are expected to grow in the near future while maintaining employee engagement and productivity, this is a world they cannot ignore.

VII. REFERENCES

[1]. Kamenov, K. Immersive Experience—The 4th Wave in Tech: Learning the Ropes. Available online: https://www.accenture.com/gb-en/blogs/blogs- mmersive-experience-wave-learning-ropes (accessed on 21 May 2021).

[2]. Friesen, N. The Textbook and the Lecture: Education in the Age of New Media; Johns Hopkins University Press: Baltimore, MD, USA,2017; ISBN 9781421424330.

[3]. Milgram, P.; Takemura, H.; Utsumi, A.; Kishino, F. Augmented reality: A class of displays on the reality-virtuality continuum. InTelemanipulator and Telepresence Technologies, Proceedings of the Photonics for Industrial Applications, Boston, MA, USA, 31 October—4November 1994; Das, H., Ed.; SPIE: Bellingham, WA, USA, 1995; Volume 2351, pp. 282–292.

[4]. Slater, M.; Sanchez-Vives, M.V. Enhancing Our Lives with Immersive Virtual Reality. Front. Robot. AI 2016,3, 74. [CrossRef]

[5]. Pellas, N.; Mystakidis, S.; Kazanidis, I. Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the lastdecade scientific literature. Virtual Real. 2021,25, 835–861. [CrossRef

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Strategic Communication (Case Study)

Vidyasagar A, Vaibhav V, Nilesh Kamath, Goutam Naidu
Department of Computer Science and Engineering, New Horizon College of Engineering Bengaluru, Karnataka, India

ABSTRACT

Space-to-Space Communications System (SSCS) is a complex two-way data communication system designed to provide voice and telemetry between three rotating systems: Space Shuttle orbiter, International Space Station (ISS), and Extra Vehicular. Activity Mobility Unit (EMU). (EMU is a space suit worn by aeronautician during space travel, or Extra Function.) SSCS is designed to allow simultaneous communication between up to five users. The program includes three product lines: space suit radios (SSER), Shuttle orbiter radio (SSOR), and Space Station radio (SSSR). All three share similarities, but have distinct features and designs.

I. INTRODUCTION

The main consideration driving the SSCS system was the construction of the Space Station. Since Space Shuttle was an existing operating system with active radios, there was no schedule pressure from the Shuttle system or EVA Project Office to deliver SSCS on a specific aircraft. The Shuttle system retained a busy launch plan in the late 1990s, and another plane was always nearby. Space Station, however, had some difficult and fast days.

From Lemke's point of view, the pressure of the schedule associated with Space Station is due to political considerations, not technical concerns. Most of the Space Station was built by a large government contractor who was late and over budget. As a result, any portion of government-sponsored goods (GFEs) arriving late, regardless of their nature, gives the contractor the opportunity to say that the government is holding the jobs and restricts its ability to deliver on time. "There was a lot of pressure to deliver on time, or everyone knew the Station was lagging behind, or everyone knew you didn't need those radios to check out the US lab module," Lemke said. "It simply came to our notice then. It was a chicken game, and the government would not hesitate to bring it."

II. DEEP SPACE COMMUNICATION

Communicating with distant spacecraft is difficult. Signs passing between spacecraft and sub-stations are very weak and because of the great distances, it takes them a long time to travel between the two. It may take up to 24 minutes for the signal to travel between Earth and Mars, for almost a day to receive a signal sent by NASA's Voyager 1 - a spacecraft that has passed the edge of the Solar System. [1]

- With powerful mechanical barriers in the space shuttle, many sophisticated communication technologies are being installed in subway stations. Numerous Discovery & Preparation

studies have contributed to the development of that technology.

Fig. 1. Deep Space Communication Antenna[1]

- A study that ended in 2012 examined the possibility of developing klystron completely within Europe. These devices convert electrical energy into magnified radio signals to send commands from sub-stations. The study set out the requirements and objectives for the development of such a facility, as well as a description of the future industrial area and road map of the future.
- Klystrons is now used in the ESA network of sub-stations; you can find out more about them in the video at the beginning of this article.
- In order to reach remote space or intra-satellite communications, optical communication becomes another interesting form of radio communication because it allows more information to be transmitted; this increases the return of science and can empower new types of machines. However, light signals are strongly influenced by Earth's atmosphere.
- Given the growing need for downlink data rate to maximize scientific retribution, the 2016 Deep Space Optical Communications Architecture Study (DOCOMAS) introduces how technology needs to change in the future to enable visual interaction between deep space probe and Earth.
- The study focuses on the lower part, which includes cloud mitigation strategies. It has identified the key technologies that allow optical ground antennas, novel photon detectors and the standard design method for the optical charging terminal. The concept design is designed for ESA's Asteroid Investigation Mission (AIM), which has evolved into Hera's mission.
- DOCMAS is based on the results of previous research focused on the development of interplanetary technology and technology, which includes investigating the necessary optical technologies. As communication is a well-known bottle in interplanetary science and exploration work, the goal was to pave the way for technological development to improve communication skills.

III. NAVIGATION TURNING TIME INTO DISTANCE

Good communication is important not only for collecting scientific and state data, but also for navigating spacecraft using the Solar System. To navigate a spacecraft we need to know where it is, which is not easy if it is very far away. But by measuring three parameters - the distance, speed and angle at which the spacecraft is located in the sky - it is possible to calculate the satellite's position on the ground in a small box-shaped space.[5]

One important factor in navigating deep space is time, especially to ensure that space time coincides with time on earth. To calculate the spacecraft's position in the Solar System, we accurately measure the time it takes for the electric current to travel between the spacecraft and the earth's orbit. Earth Navigators then transmitted the subject changes. A 2007-2009 study explored forward thinking strategies to adjust the duration of a deep space probe board to get accurate

navigation, especially looking at less expensive options. Coherent studies have found that the precision of ten nanoseconds of signal passing through Earth's orbit would be possible without the use of an atomic clock.

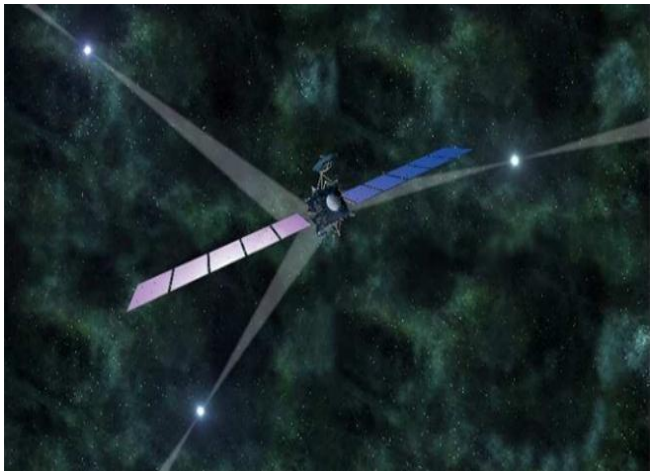


Fig.2. Star Navigation [1]

Providing communication service across the globe during unforeseen circumstances have been a major problem in our decade. When we are planning missions to the moon, mars & beyond, we need more sophisticated network laid across solar system for effective communication. During the Solar Conjunction of 2021-22, there was zero communication between all the mars satellite & the Earth. If we are sending people to mars, then we need at least minimal communication between earth & mars.

Using various strategies & placing Amplification communication satellites across the solar system will improve communication & helps the mission better.

IV. COMMUNICATION STRATEGY

Mars Communication Strategy [6]

- This constellation contains 3 customized amplification satellites placed in a solar synchronous orbit in the initial stages of testing and research.
- Satellites placed in these orbits provide communication at all times towards earth and also with one another.
- Since the satellites are placed between earth and mars orbit, there is no worries about meteors as the asteroid belt is between mars and Jupiter.
- There will be one more sat orbiting in the high orbit of mars which will be the main transmission for communication constellation.

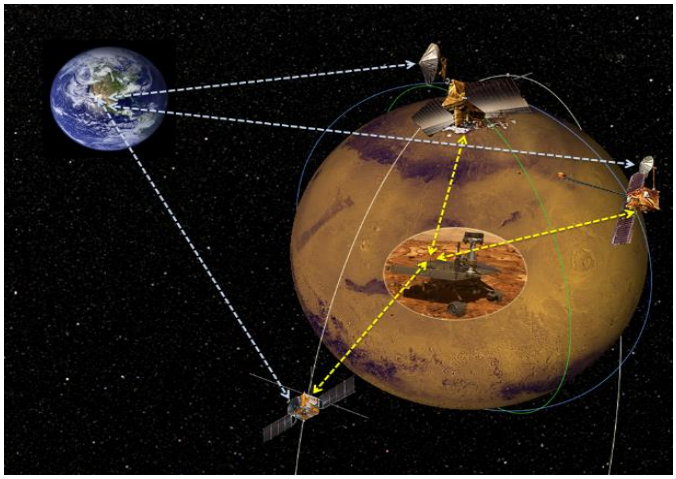


Fig. 3. Mars Communication Strategy[2]

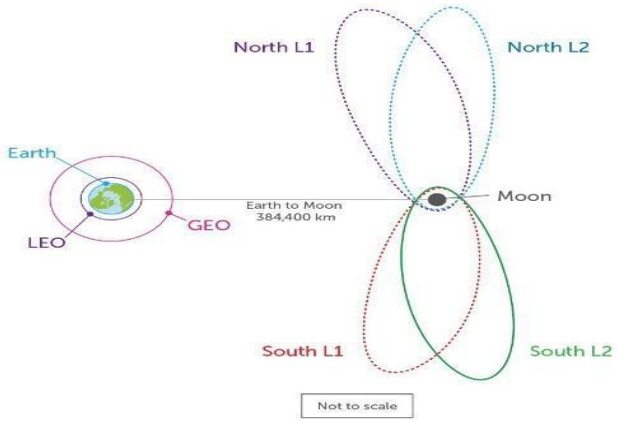


Fig. 4. Moon Communication Strategy[3]

- Just placing the already existing orbiters in the right orbit can give proper and better communication.
- Consider a circular planar orbit around moon such that one plane of orbit always faces earth. In this case continuous and qualitative communication can be maintained at all times.

Redundant antennas for critical systems to be installed.

- This communication strategy is under research and will be first demonstrated on the Artemis mission which will launch in May 2022.

V. REFERENCES

[1]. ESA - Deep space communication and navigation
[2]. http://www.nasa.gov/press/2014/july/nasa-seeks-proposals-for-commercial-mars-data-relay-satellites
[3]. https://blog.maxar.com/space-infrastructure/2019/what-is-cislunar-space-and-a-near-rectilinear-halo-orbit
[4]. NASA Artemis & ESA - Navigation
[5]. Relay communications strategies for Mars exploration through 2020 (researchgate.net)

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Newspaper Analysis using Data Mining and Machine Learning

Kishore Kumar L[1], Lokesh B S[1], Kumar S[1], Nandan KN[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore-560103, Karnataka, India

ABSTRACT

Text categorization is one of the most important applications of data mining. Text classification separates text documents into class labels based on words, phrases, and word combinations. The current study divides news data into four categories: Times Now, The Hindustan, The Times of India, and The Hindu. WEKA, an open-source data mining tool, is used for text classification. On a data set of news articles, various classification methods are performed, and a comparison of some machine learning algorithms is conducted.
Findings: Predicting the newspaper brand using data mining and machine language concept.

Keywords: WEKA, Machine learning, Classification algorithms, Data mining, Text classification

I. INTRODUCTION

Because of the ease with which they may be accessed, the number of digital documents is rapidly expanding in this digital era.. This is made feasible by high-capacity hardware, software, and powerful computer and storage devices, all of which are readily available at low prices.
[1]Data mining can be used to extract or find new relationships, hidden knowledge, and significant patterns from large amounts of data. Knowledge Discovery in Databases is another name for data mining (KDD). Data mining uses different technique for knowledge discovery such as classification, clustering, summarization, associations etc. Text mining is one of technique used in data mining for analysis of large volume of textual data. Analysis helps in discovering useful patterns and extracts information. Text classification is one of the key technologies used in text mining. It is useful in various applications such as automatic indexing of research and scientific articles, tracking customers email for Customer Relationship Management (CRN), categorizing web documents, detection and identification of terrorist activity etc... It is useful for document filtering, information extraction, machine translation, language identification and document classification[1]. Vast availability of any type of documents needs its effective organization and retrieval. To have easy access to these documents these documents are classified using text classification techniques. It is one of the important techniques for categorization of documents in supervised way.
[3]A news article discusses current or recent news of either general interest (i.e., daily newspapers) or on a specific topic (i.e., political or trade news magazines, club newsletters, or technology news websites). A news article can include accounts of eyewitnesses to

the happening event. We must have seen the news divided into categories when we go to a news website. Some of the popular newspaper categories that you'll see on internet are Times Now, The Hindustan, Times of India and The Hindu.[3]

[14]Every newspaper has their own unique style of words and articles. An article's style of words and sentences in a newspaper is different from each other. For instance, the rhythm of words in Times of India is different from The Hindu and so on. So, by collecting the Dataset of major newspaper articles, we can classify the newspaper brand by applying Text classification and then machine learning concept. Thus, the system will be able to predict the newspaper's name (output) by analyzing the article (input).[14]

Text Classification

Text classification datasets are used to categorize natural language texts according to content. For example, think classifying news articles by topic, or classifying book reviews based on a positive or negative response. Text classification is also helpful for language detection, organizing customer feedback, and fraud detection.

While this process is time-consuming when done manually, it can be automated with machine learning models.

For news, categorization is a multi-label text classification issue. The purpose is to allocate a news story to one or more categories. The employment of a set of binary classifiers is a common strategy in multi-label text classification.

We have gathered a dataset of articles from prominent newspaper brands in India, such as Times Now, The Hindustan, Times Of India, and The Hindu, for the task of news classification using machine learning. The dataset contains news items, including headlines and categories.

Data Fields

- Article Id – Article id (unique number) given to the record
- Article – Text of the header and article
- Category – Contains the newspaper name from which the article is taken

WEKA-data Mining tool

[1]The University of Waikato in New Zealand developed WEKA, an open source data mining tool. This tool's created using the Java programming language. This tool consists of number of data mining algorithms and different filters for data pre-processing. For simplicity and ease of use, these algorithms are grouped on the basis of rules generated by algorithm. This tool is very handy and simple to use and also it is freely available on the internet. Due to these reasons, we use WEKA for present study. In present study we used different classification algorithms for Text classification.

Data Cleaning and Data Pre-processing

The process of converting raw data into a comprehensible format is known as data pre-processing. We can't work with raw data, thus this is a key stage in data mining. Before using machine learning or data mining methods, make sure the data is of good quality.

Text Classification Process

Text classification was performed using a collection of predefined class labels and words, phrases, and word combinations. There is a training phase and a testing phase in the text classification process. The dataset is loaded during the training phase, and several classification methods are used to it. After the training phase is completed, the performance of the classifiers is evaluated, and the classifier with the best performance is chosen. This can be chosen on the

basis of different factors such as accuracy, time and error rate etc.[1]

II. RESULTS AND DISCUSSION

Present study uses WEKA's classification algorithm namely J48, NaïveBayesMultinomial, MultiClass Classifier (Logistic), SMO, HyperPipes and JRip for text classification purpose. Experimental result of said algorithms are summarizing in following tables. Present study use "10-fold cross-validation" as a Test mode for all algorithms and "full training set" as classifier model. The dataset is partitioned into 10 folds in the 10-fold classification method. Each fold serves as a test dataset, while the remaining nine serve as training datasets. In each fold and test set, performance is measure for different case-based configuration. Table 1 shows result based on accuracy and time taken to build model for each algorithm. Figure no. 1 illustrates a bar plot showing the average number of tokens in each newspaper channel. Figure no. 2 shows a bar plot illustrating the percentage of newspaper distribution by channel.

Predict News Article

Now, after the completion of model analysis, we can predict the newspaper name using Machine Learning.

TABLE I: TABLE OF ACCURACY AND TIME

Algorithm	Correct Instance	Percentage of Correct Instance	Time required to Build Model
NaiveBayesMultinomial	448	92.7536	0.03
J48	364	75.3623	26.46
JRip	360	74.5342	66.92
Multiclass Classifier (Logistic)	443	91.7184	649.59
SMO	434	89.8551	0.46
Hyper Pipes	447	92.5466	0.14

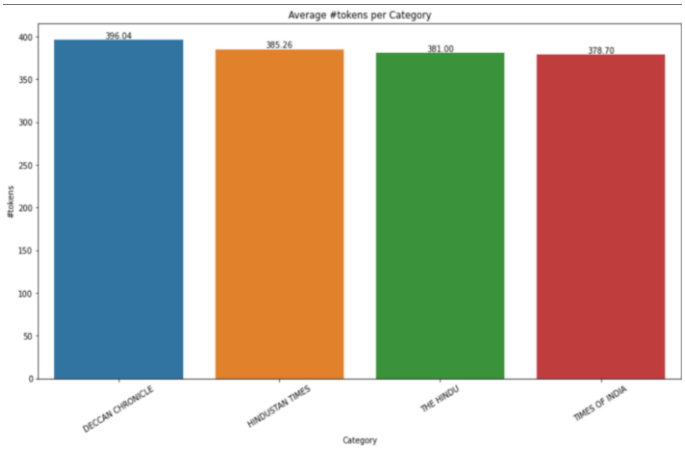


Fig. 1: A bar plot illustrating the average number of tokens in each newspaper channel.

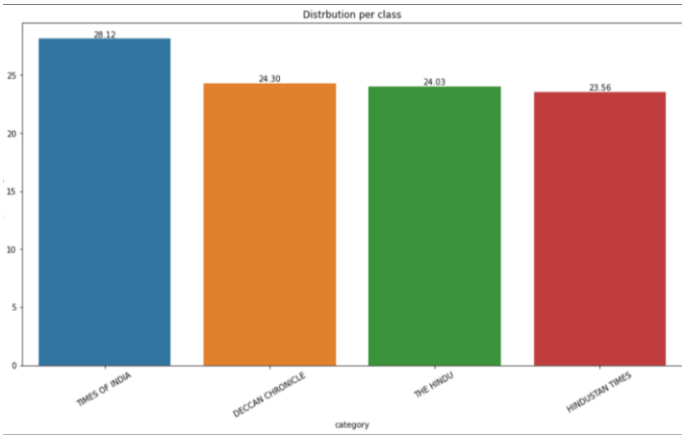


Fig. 2: The percentage of distribution of each newspaper channel is depicted as a bar plot.

III. OBSERVATION

Kappa statistics can be used to distinguish between data collection reliability and data validity. It's utilised to get to certain measuring cases11. The selected algorithms' average Kappa score is roughly 0.812033. The accuracy of this classification is substantial, according to Kappa Statistic11. For the NaiveBayesMultinomial algorithm, the highest value of Kappa statistics is 0.9023, which is close to 1.

According to Kappa statistics, the NaiveBayesMultinomial method is the best for the current news dataset.

The current study demonstrates that the NaiveBayesMultinomial method is effective for news classification. The NaiveBayesMultinomial algorithm has the highest accuracy, the shortest time and errors, and the highest ROC value.

IV. CONCLUSION

Text classification is one of the most common applications of data mining. The news in this study is classified using a text classification technique. This news dataset was compiled from the websites of the Times of India, Indian Express, and The Hindu newspapers. There are four sorts of news in this dataset: business, entertainment, technology, and sports. The current study employs the free source WEKA data mining technology for news classification. J48, NaiveBayes Multinomial, Multiclass Classifier (Logistic), SMO, and Hyper Pipes, as well as JRip algorithms from WEKA, are employed in this study. In this work we apply these algorithms on news dataset. These datasets are in text form. For experimental purpose, the present study uses WEKA data mining tool. Experimental result is analysed on the basis of accuracy. The NaiveBayesMultinomial method is the best for news classification, according to the current research. The accuracy of NaiveBayes Multinomial is the greatest, at 92.7536 in 0.03 seconds. In addition, the Kappa statistic value for this approach is close to 1.

From figure1, the Category versus #Tokens plot gives equal distribution bar for all the newspapers taken. This indicates that there no much difference in average number of tokens in each newspaper article.

From figure2, the Category versus Words depicts that the average number of words per article in Times of India is comparatively higher than other categories even though average number of tokens is same.

Finally, after doing Data cleaning and Data Pre-processing (cleaning data, train_test_split model, creating a bag of words NLP model, and machine learning model) we got the accuracy scores and we can say that NaiveBayesMultinomial gives the best accuracy among all machine learning models.

V. REFERENCES

[1]. Text Classification of News Articles Published by Analytics Vidhya

[2]. Text Classification of News Articles, DEEPAK SINGH.

[3]. News Classification: A Data Mining Approach Article in Indian Journal of Science and Technology · December 2016

[4]. Han J, Kamber M. Data mining - concepts and techniques. 3rd edition, India

[5]. Sebastiani F. Machine learning in automated text cat-egorization.Journal ACM Computing Surveys. 2002 Mar; 34(1):1–47.

[6]. Kawade DR, Oza KS. SMS spam classification using WEKA. International Journal of Electronics Communication and Computer Technology. 2015 May; 2015. page. 5.

[7]. Delany, SJ, Cunningham P. An analysis of case-base edit-ing in a spam filtering system. Advances in Case-based Reasoning. Springer Berlin Heidelberg; 2004. page. 128–41.

[8]. Li-gong Y, Jian Z, Shi-ping T. Keywords extraction based on text classification. Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM-13); 2013.

[9]. WEKA [Internet]. [cited 2015 Aug 21]. Availablefrom: http://www.cs.waikato.ac.nz/~ml/weka.Junaid

MB, Farooq M. Using evolutionary learning classifiers to do mobile spam (SMS) filtering. GECCO'11, Dublin, Ireland; 2011 Jul.

[10]. Kappa [Internet]. [cited 2015 Oct 12]. Available from: http://www.dmi.columbia.edu/homepages/chuangj/kappa.

[11]. Available from: http://gim.unmc.edu/dxtests/roc3.htm.

[12]. Available from: https://en.wikipedia.org/wiki/Receiver_operating_characteristic.

[13]. Kang MY, Shin J-D, Kim B. Automatic subject classification of korean journals based on KSCD.

[14]. News Classification: A Data Mining Approach Article in Indian Journal of Science and Technology · December 2016

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Market Segmentation Using Cluster Analysis

Akash Kumar Barik, Aditya Raj, Avni Garg, Aaqib Nissar
Department of Computer Science, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Customer segmentation enables marketers to achieve a more profound understanding of customers likes and dislikes, it is a more effective methodology followed by marketers all over the world. In order for businesses to stay in the game for a longer run, they mainly need to figure out which segment of the market they want to serve. This is also called target marketing, it works on dividing the market into several segments and develops products or services for those particular segments, that try to keep up with the customer needs. Machine learning algorithm find statistical regularities in data, which can process customer data and patterns and help analyze the best suited segment or clusters, which is very difficult to spot through intuition and manual examination of data. This study aims to provide a better understanding of how clustering algorithm helps to better understand customers with respect to both static demographics and dynamic behaviors and suggesting improvements. Keywords: Machine Learning Algorithms, Market Segmentation, Target Marketing, Clustering

I. INTRODUCTION

Successful market segmentation leads to a healthy and powerful Customer Relationship Management (CRM). The main objective is to build stable relationships and to optimize the financial outcomes of a company. Companies that have managed to adapt their offering to one or more segments of the markets have gained an advantage over the alternatives of mass marketing or mass producing [1]. As companies collect more data and more detailed data about customers, the advantages of targeting specific groups of homogeneous users compared to using a mass marketing approach will continue to grow [2].

Since the marketer's goal is usually to maximize the value (revenue and/or profit) from each customer, it is critical to know in advance how any particular marketing action will influence the customer. Ideally, such "action-centric" customer segmentation will not focus on the short-term value of a marketing action, but rather the long-term customer lifetime value (CLV) impact that such a marketing action will have. Thus, it is necessary to group, or segment, customers according to their CLV [3].

For instance, when we look at customer data, we may have data from millions of customers, these customers may belong to a few selective segments: customers are kindred within each segment but across numerous different segments. Frequently we may feel the need toanalyze each segment separately, due to their behavioral difference (e.g. different market segments may have different product or service preferences and behavioral patterns).

In such situations, to identify segments in the data one can use statistical techniques broadly called Clustering techniques. Based on how we define

"similarities" and "differences" between data observations (e.g., customers or assets), which can also be defined mathematically using distance metrics, one can find different segmentation solutions. A key ingredient of clustering and segmentation is exactly the definition of these distance metrics (between observations), which need to be defined creatively based on contextual knowledge and not only using "black box" mathematical equations and techniques [4].

The "Business Decision", Management of a large corporation would be invested in understanding the types of people who are, or could be their possible clients, interested in their products or services.They have good reasons to believe that there are a few different market segments, and they are considering designing and positioning the corporation services better in order to attract mainly a few profitable market segments, or to differentiate their services (e.g. invitations to events, discounts, etc.) across market segments [4]. It offers an opportunity to pinpoint exactly what messaging will drive your customers to make a purchase.

The four basic types of customer segmentation are:

1. Demographic: technique where a firm's target market is segmented based on demographic variables such as age, gender, education, income, etc.
2. Psychographic: methodology used for studying consumers and dividing them into groups using psychological characteristics including personality, lifestyle, social status, activities, interests, opinions, and attitudes.
3. Geographic: segmenting your audience/customers based on the region they live or work in. This can be done in numerous ways i.e. (grouping customers by the country they live in, or smaller geographical divisions, from region to city, and right down to postal code).
4. Behavioral: segments depending on their behavior patterns when interacting with a particular business or website.

II. METHODS AND MATERIAL

There are various solutions available in the market to solve customer/market segmentation, from packaged software to CRM products. This paper solely depends on how clustering algorithms can help start-up companies to choose or target their appropriate segment. Machine learning algorithms are divided into two parts:

1. Supervised
2. Unsupervised

Here, we are going to use unsupervised learning algorithm, called K-means Cluster Algorithm. The data-set we are using is from November 2018 — April 2019, it is an actual sales data courtesy of an e-commerce company.

K-means clustering will be applied via the following methods:

I. Business case Study
II. Preparing the data for analysis
III. Applying K-means for segmentation
IV. Hyper parameter Tuning
V. Visualizing the end result

2.1 Business Case Study:

In this paper, we visualize the customer behavior, interaction and characteristics from various aspects. Taking it one step further, we will form the business study around the question: Can the customer base be grouped to develop customized relationships? The approach made for this question will be from a behavioral aspect (alternatives can be geographical or demo-graphical perspectives) to better understand customer's spending and ordering habits with the

following features: Number of products ordered, average return rate and total spending.

2.2 Data Preparation:

Data-set is well-formatted and has no NA values. So, we can start by forming the features. 3 features will be calculated per customer id and they will help us with the visualization (using Plotly library) and algorithm explain-ability in the latter steps. Data preparation will be done with pandas and numpy.

Scaling:

K-means algorithm elucidate every row in the customers data frame as a point in a 3-D space. To group them together, the Euclidean distance is taken in account which is the distance between the data points and the center of the group. When the highly varying ranges occur, the algorithm may perform deficiently and it will not make groups as expected. To perform the algorithm effectively, we are going to scale the data using logarithmic transformation which will be more suitable transformation for skewed data. This algorithm will scale down the 3D space in which our data is spread, yet it will maintain the proximity between the points [6].

2.3 Applying K-means Clustering:

This is how the algorithm will form customer groups:

We will Initialize k=n centroids=number of clusters randomly or we will assign each data point which is the closest centroid based on Euclidean distance, and thus the group will be formed. Now we will move centres to the common average of all points in the given cluster. We will then Repeat steps until convergence. While running through the steps, the algorithm will check all the sum of distances between clustered point and centre for each cluster. Mathematically, it tries to minimize and optimize within the cluster sum distances or inertia of each cluster. This algorithm is easy to understand and it fits well to large datasets in terms of computing times and it also guarantees convergence.

However, when centroids are initialized randomly, algorithm may not assign the points to the groups in the most optimal way. The one more important consideration is the selection of k. In the next step, we are going to choose k.

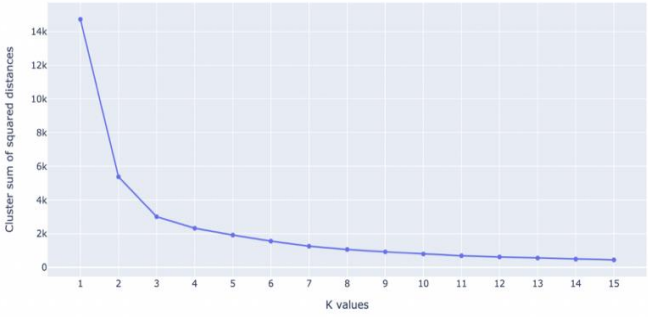


Figure 1. Elbow method

2.4 Hyper-parameter Tuning:

While selecting k, we are going to decide against the optimization criteria of the K-means, inertia, using elbow method. We are going to build different K-means models with k values 1 to 15, and save the corresponding inertia values. With the elbow method, we are going to select the k value where the decrease in the inertia stabilizes.

When k=1 inertia is at the highest, meaning data is not grouped yet. Inertia decreases steeply until k=2. Between k=2 and 4, the curve continues to decrease fast.

At k=4, the descent stabilizes and continues linearly afterwards, forming an elbow at k=4. This points out the optimal number of customer group is 4 [5].

2.5 Visualization and Interpretation:

Data points are shown in spheres and centroids of each group are shown with cubes. The four customer groups are taken as follows:

i. Blue: Which shows the Customers who have ordered at least 1 product and with maximum spending of 100 and having the highest average

return rate. They might be the newcomers of the e-commerce website.

ii. Red: These are the customers who have ordered 1 to 4 products, with average spending of 150 and a maximum return rate of 1/2(0.5).

iii. Purple: These are the following Customers who have ordered 1 to 4 products, with average spending of 300 and a maximum return rate of 1/2(0.5).

iv. Green: These is the category of Customers who have ordered 1 to 13 products, with average spending of 600 and average return rate as 0. Which is why It makes the most favorable customer group for the company.

Now we will see the cluster magnitudes or the number of customers in each group. The Blue group is the 42% of all customers, any improvements in this group will increase or change the revenue.

To Eliminating high return rates and by giving gift cards can move the customer group to low average return rate and push them to high total spending area.

Now let's assume that they are some newcomers, gift cards can accelerate their come-back. Red and purple group together consists of 50% of all customers. They are showing the same characteristics from the average return rate and products ordered perspectives but they differ from total spending. These groups can be categorized as who already know the brand and orders the multiple products. Those customers can be kept up-to-date with the brand with some specialized categories according to the convenience. Green customer group consists of 8% of all customers, resulting in becoming the most favorable customer group for the brand.

They order multiple products and they are highly likely to keep them. To maintain and possibly expand of this kind of groups, special deals and pre-product launches might help. Moreover, they can be magnets to attract new customers and exert the influence in expansion of the customer base.

III. RESULTS AND DISCUSSION

Applying the above methodology, we have understood that market data is always very huge and analyzing it using various data mining algorithms for maximizing the profit is something very vital. The dataset that we have used here consists of weekly sales and product number which is named as dept. In this research we make sure to predict the behavior of customers with the given data and from the above analysis we have understood that k means can be used for analyzing the customer behavior according to the market. K-means clustering is useful since it can work well with distinct points as well as large dataset. It is applied on the top profitable as well as correlated products which helps us to maximize the market's profit. Figure 2 and Figure 3 shows the results of the above analysis. Since our dataset is huge and lacks the customer information, k means helps us to analyze accordingly [6].

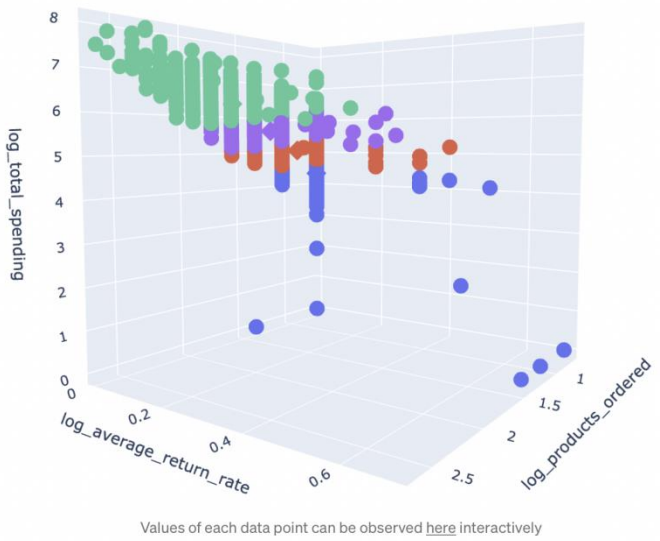


Figure 2. 3-D representation of the Clusters formed

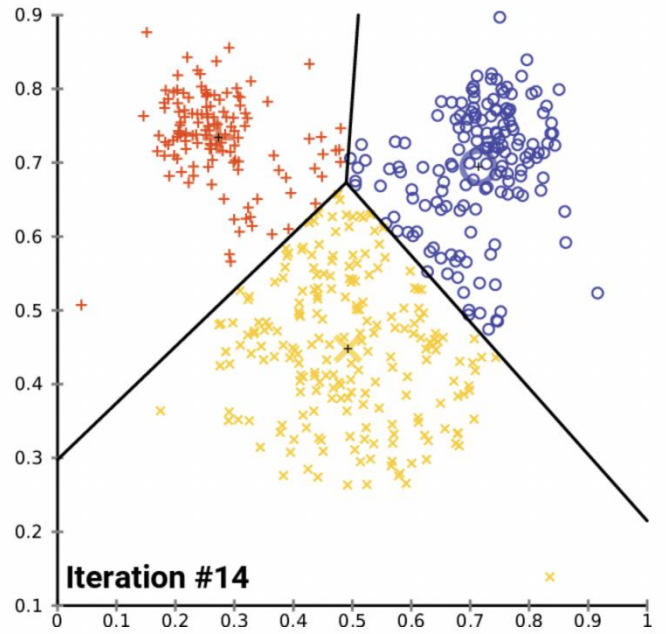


Figure 3. Formation of Cluster after 14 Iterations.

IV. CONCLUSION

We have approached the customer segmentation problems from a behavioral aspect with the number of products ordered, average return rate and total spending for each customer. We have understood and visualized the model with the help of 3 features.

Overall, the data-set was suitable to perform an unsupervised machine learning problem. In the starting, we only had customers data without any categorization and without any information about the group in which the different data will be grouped together.

With the help of K-means clustering, patterns and the tracing in the data were found and it was extended and categorized further into groups. We have carved out strategies for the formed groups out of a dataset that was a dust cloud initially.

With the growing population and demand, it was very important for the markets to extract and bring out the patterns of their product sales and understand their customer's behavior to profitably run their business by meeting their customer demands at same time.

Nowadays, Data analysis techniques have become a crucial tool in the market industry to obtain patterns from the available data and predict the future of the market and make profitable decisions accordingly.

It is also important for the business owners to forecast the future of the market with the obtained data in order to maximize their profitability and satisfy their customers in long competitive run.

We have used k means clustering model for clustering the data-set and obtaining various inference with reference to the sales of every profitable product. This approach helps the market to maximize their profitability and ensure the fulfillment of demand.

V. REFERENCES

[1]. Kamthania, Deepali, Ashish Pawa, and Srijit S. Madhavan. "Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business." Journal of computing and information technology 26.1 (2018): 57-68.

[2]. Thomas, Jerry W. "Market segmentation." Decision Analyst,Arlington, Texas, USA (2007).

[3]. Florez-Lopez, Raquel, and Juan Manuel Ramon-Jeronimo. "Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting." Social Science Computer Review 27.1 (2009): 96-117.

[4]. Michel Wedel and Wagner A Kamakura. Market segmentation: Conceptual and methodological foundations, volume 8. Springer Science & Business Media, 2000.

[5]. AmbjornKarlsson,Market Segmentation with Cluster Analysis based on Video Streaming Data , Degree project in computer science and

engineering, second cycle, 30 credits Stockholm, Sweden 2017.

[6]. Sruthi Janardhanan and Raja Muthalagu , "Market segmentation for profit maximization using machine learning algorithms", 2020 J. Phys.: Conf. Ser 1706 012160.

[7]. Hung, Phan Duy, Nguyen Duc Ngoc, and Tran Duc Hanh. "K-means clustering using RA case study of market segmentation." Proceedings of the 2019 5th International Conference on E-Business and Applications. 2019.

[8]. Tuma, Michael, and Reinhold Decker. "Finite mixture models in market segmentation: a review and suggestions for best practices." Electronic Journal of Business Research Methods11.1 (2013): pp2-15.

[9]. Lin, Jihao, et al. "A Machine Learning Approach to Online Market Segmentation." Available at SSRN 3941093 (2021).

[10]. Liu, Ying, et al. "Multicriterion market segmentation: a new model, implementation, and evaluation." Marketing Science29.5 (2010): 880-894.

[11]. Nilashi, Mehrbakhsh, et al. "Online Reviews Analysis for Customer Segmentation through Dimensionality Reduction and Deep Learning Techniques." Arabian Journal for Science and Engineering 46.9 (2021): 8697-8709.

[12]. Wang, Chih-Hsuan. "Outlier identification and market segmentation using kernel-based clustering techniques." Expert Systems with Applications 36.2 (2009): 3744-3750

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Role of Humans in Machine Learning

Guna Keerthi. P, Kavya. S, Gangireddy Ramya Sri, Guggulla Geethapriyanka
Department of Computer Science and Engineering, New Horizon College of Engineering Bangalore, Karnataka, India

ABSTRACT

Machines are an important part of our daily lives, and we interact with them effectively. Humans and machines both learn, which develops our technology and expands the applications of machines in the human world. Several case examples are presented to demonstrate how machines learn from humans. In complicated strategic situations, we investigate how human and machine learning vary. Despite their amazing effectiveness, modern machine learning techniques do not always overcome enough light on human learning.

Keywords—Bioinformatics, Distilling principles, Gesture-based music, machine learning, Reinforcement learning, Transparency

I. INTRODUCTION

Machine with the rapid expansion in the storage capacity and processing power of computers, the discipline of machine learning, which may be roughly defined as enabling computers to make effective predictions based on past experiences, has seen impressive growth recently. Machine learning approaches have been widely used in bioinformatics and many other areas. Humans and machines have a lot in common when it comes to learning. How people learn varies from one person to the next. It is difficult to change a learning process after it has been embedded in people's thoughts. However, in Machine Learning (ML), changing the learning method is as simple as selecting a different algorithm. We have well-defined techniques in Machine Learning to analyze and measure learning accuracy. Human learning is typically assessed by examinations, and it is not considered a measure of IQ.

II. INTERACTIVE MACHINE LEARNING

1. Preface

Designing and using algorithms and intelligent user interaction frameworks that promote machine learning with the help of human participation is known as interactive machine learning. By communicating with people in the natural language and awareness we want computers to learn from them. Computers can also be taught by humans

2. Interactive Machine learning in image processing

Graphic processing is a very important technology, and the demand for industry seems to be growing every year. Digital image processing began to emerge in 1960's as a way to duplicate the human visual

system and to perform the process of automatic image analysis. Solutions for specific task begin to emerge as technology progress and developed.

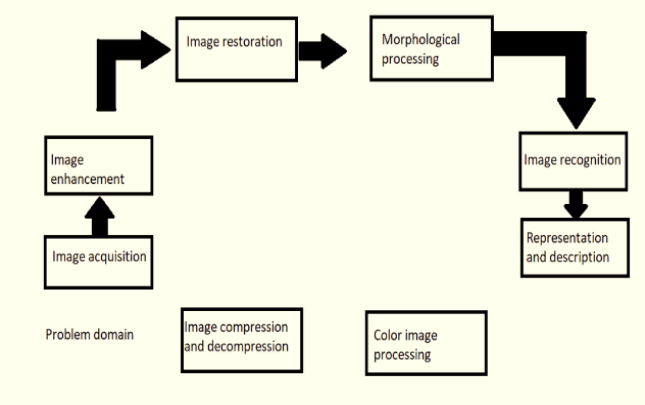


Fig 1: Image processing

3. Interactive Machine Learning with touch-based music

Touch recognition is a system that learns and interrupts hand gestures as directed through the senses. This capability allows drivers and passengers to interact with the vehicle-usually controlling the infotainment system without touching buttons or screens in the automotive industry. When artists associate with musical instruments, for example, they tend to get a quick response.

Fig 2: Touch-based music

III. USERS ARE PEOPLE, NOT PREDICTIONS

Engineers are the once who usually put "Exercised Learning" into action. They make sure that technology is strong enough to be integrated and used in the most important mechanical systems. Certain aspects of machine learning programs, including possible reduction and flow of project times, may need to notify to supervisors.

Data Engineers have been charged with building data infrastructure for various other applications, such as business intelligence, and their skills will be required to adopt a learning machine. Data scientists are in charge of coming up with data-driven solutions to the problem of the commercial world. They may, for example, analyze user data to identify key user segments and create models that can differentiate those people into segments to differentiate end user experience and increase engagement. Engineers who specialize in machine learning have emerged as ML production specialists. To summarize this approach, data scientists construct and validate the model, while engineers ensure that the model is analyzed from the prototype to the production system.

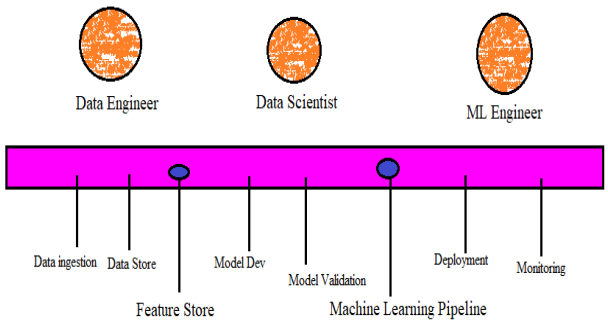


Fig 3: Role of Engineers

1. People tend to give student better feedback than bad

Enhanced learning is a machine learning approach developed by a behavioural scientist, according to Techopedia, a source of expertise to the acquisition of professional IT knowledge and inspiration. It is similar with the way a child learns to do something new. Strengthening learning is different from other machine learning methods because the algorithm does not tell you how to complete the task clearly, but

instead solves the problem itself. [1] Thomas and Breazeal (2008) noted that people have a stronger tendency to give positive rewards. [2] Knox and Stone (2012) later confirmed this positive bias in their experiment.

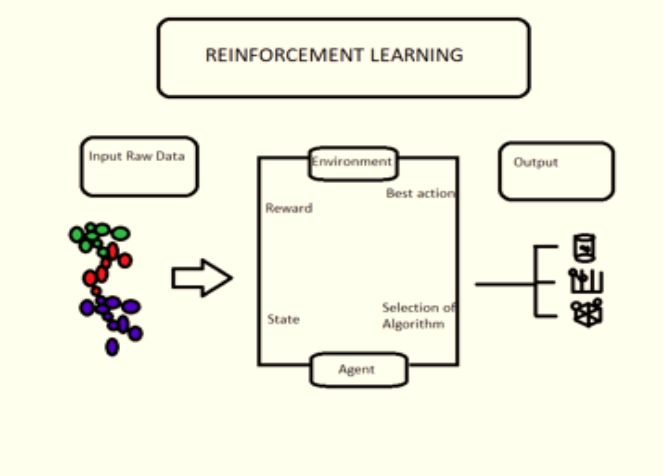


Fig 4: Reinforcement Learning

2. The role of people in demonstrating student style in machine learning

[3] Although Reinforcement Learning (RL) is not usually built into interactive monitoring input from a human instructor, various robotics courses and software agents have adapted human input by allowing the human trainer to control the reward signal. We present three important details about how people give feedback to a robot when they use Reinforcement Learning to teach a job: (a) Users of a rewarding channel not only for feedback but also for future-oriented instructions; (b) they show good bias in their response – perhaps as an incentive; (c) They change their behaviour as they create a robotic student mental model.

3. Naturally people want to provide Data Labels Only

[4] Labelling machine learning data such as making high-quality data sets for AI model training. If the model is based on a visual vision model, then computer-based training data will usually be available in image or video format. Label data remains the most popular way to integrate end user into machine learning systems due to its simplicity and ease of use.

4. People are very important openness in learning Programs

Principles of public machine learning Transparency is a feature of the app in general. It's about much internal system performance can be understood "in theory." It can also point to a way to deliver easy-to-use descriptions of models and algorithmic decisions. Smartphones and sensors can now provide information about a user's work and even his or her physiology due to the development of machine learning. However in addition to directing the recovery process, the collection and transfer of IoT - data are also prone to encryption. [5]The most common question for anyone trying to choose an environmentally friendly fashion is: who created my clothes? How are they made? It comes second. Although it may not seem like hard work, it is often difficult to find information. It's not as easy as looking at a label; instead, it requires checking the product website or contacting them via email or social media.

IV. CONVERSATION

[6] Collaborative machine learning is a way to allow the end user's participation in the most powerful machine learning. Studying how people interact with interactive machine learning systems and exploring new strategies to help that interactive machine learning systems and exploring new strategies to help that interaction can lead to improved self-awareness of the most effective user and machine learners, as this article shows. However, research in this area is still in its infancy, and there are still many ways to improve the process of machine learning. This section

discusses the current state of the art in human communication and interactive machine learning systems, as well as the problems and opportunities that exist.

1. Lock principles and guidelines for designing personal Interaction with machine learning

[7] There is still space for productivity from existing solutions, as well as dissipation principles and guidelines on how to design future human interactions using interactive machine learning, as we have in conventional communications (e.g., Schneiderman). Et al., 2009; Moggridge and Smith, 2007, Dix et al., 2004, Winograd, 1996; Norman, 1988) Machine learning systems provide users with feedback about their behaviors, and giving users more control over machine learning systems provide users with feedback about their behaviors, and giving users more control over machine learning systems can often improve user experience, as this article explains. Interactive machine learning systems, on the other hand, frequently run afoul of any existing visual connector design rules. Traditional linkages that promote comprehension (i.e., predictable or clear programmers regarding how they work) and performance (i.e., programmers that make it clear how to achieve one's goals while allowing one the freedom to do so) According to research, are typically more useful than networks that do not support these goals. Many machine learning systems break both of these principles: they are inherently difficult for users to comprehend and drastically limit the possible to investigate how modern design concepts apply to human-computers interactions in machine learning. [8] Norman (1994); Hook (2000); Horvitz(1999), There are only a few researchers who have come up with novel approaches to developing intelligent software. When developing intelligent systems, Norman (1994) and Hook (2000), both highlight security and trust as crucial factors, alluding to the anti-certainty of preventing undesirable adaptation or behavior.

2. Algorithmic problems in interactive machine learning

Two significant technical challenges exist when using an interactive machine to learn user interaction. To begin with, the necessity for rapid model updates frequently necessitates a trade-off between accuracy and speed. As a result, new, model are inaccurate. [9] While more interactive machine learning duplication can help solve this problem, faster and more accurate methods can improve the quality of learned models and cut down on the number of repeats needed to get relevant models. [10] Second, as many of the research in this article have shown, users may desire to interact with machine learning systems in ways that present machine learning approaches do not support. A new framework and algorithms that can manage a variety of desired and natural traits can be developed to overcome this problem.

V. CONCLUSION

The three principles in this easy are backed by the instances of teachings offered. To begin with, interactive machine learning cycles are more focused, faster and more advanced than traditional machine learning cycles. [11] This increases the chances of users influencing the users. As a result, the system's features and user output are inextricably intertwined, necessitating a strong requirement to investigate the system closer to possible users. Second, in order to improve this decision, it is required to examine the users of the learning programs. Ongoing user research can aid in the identification of user requirements and desires, as well as inspire new user interactions with the product. [12] With the introduction of new

communications, user studies can also help uncover common barriers that users face. Finally, there is no reason to limit communication between learning programs and their users. We can create robust interactive machine learning systems giving end users more control than just capturing situations, and by providing more light to users than just the expected output of the reader. More users control and more openness from the reader, on the reader.

VI.REFERENCES

[1]. Power to the people: the role of humans in interactive machine learning. - Free Online Library (thefreelibrary.com)

[2]. Human engagement providing evaluative and informative advice for interactive reinforcement learning | SpringerLink

[3]. Creating Interactive Crowds with Reinforcement Learning (aaai.org)

[4]. Machine Learning: Algorithms, Real-World Applications, and Research Directions - PMC (nih.gov)

[5]. Machine learning - Wikipedia

[6]. Interactive machine learning - Google Scholar

[7]. Guidelines for human-AI interaction design - Microsoft Research

[8]. NEPAPER\LWA09\LWA09HARTMANN.TEX (psu.edu)

[9]. Ph.D. Dissertation. University of Washington, Seattle, WA. Amershi, S., Cakmak, M., Knox, W. B., Kulesza, T., & Lau, T. 2013. IUI workshop on interactive system learning. In Proceedings of the 2013 International Conference on Intelligent User Interfaces Companion (pp. 121-124).

[10].ACM. Blackwell, A. F. 2002. First steps in programming: A cause for interest funding models. In Human Centric Computing Languages and Environments, 2002. Proceedings. IEEE 2002 Symposia on (pp. 2-10). IEEE.

[11].Cakmak, M., Chao, C., & Thomaz, A. L. 2010. Designing interactions for robotic energetic learners.

[12].Autonomous Mental Development, IEEE Transactions on, 2(2), 108-118.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Analysis of Use of Big Data in Google

Karthikeyan Umesh, K V Charan Reddy, J Karthik Surya
Department of CSE, New Horizon College of Engineering Bengaluru, Karnataka, India

ABSTRACT

Google's power is derived almost entirely from how it uses the billions of pieces of information it collects on its users—a collection of information known as big data. This paper analyses big data use cases in Google. It discusses the different ways it collects data from users and identifies the data tasks of Google to process it and applies different ML algorithms on it to use the big data for its personalization services.

Keywords—Big Data, Google, Big Query, Google Cloud Dataflow, Cloud Dataproc, Pub/Sub, Data Fusion, Data Composer, Data Catalog, Data Studio, Data Transfer, Cloud Bigtable, Cloud Datalab

I. INTRODUCTION

Google uses big data to understand how big data is expected based on multiple parameters such as search history, location, and trends. Then run the algorithm to make a complex guess. Google displays an ordered list of indexed indexes in terms of importance and authority and is ranked. This is what users expect. Google makes it easy to view results based on the relevance and permissions that suit your needs. Google uses a variety of techniques such as indexed pages, real-time feeds, sorting tools, knowledge graphs, Semantics Search, and Google Translate. Google has a huge impact on how we analyze large datasets (MapReduce, BigQuery, etc.), but making it a part of our daily lives is more responsible than anyone else. Most Google innovations today will be implemented by most companies in the coming years. Today's massive data preparation goes far beyond simple searches, but it's still their core business. They process 3.5 billion requests per day, and each request queries the database of 2 billion websites. Originally PageRank contains information about websites that link to a particular website and help you assess the overall importance of that website. So far, major search engines have been working almost exclusively on the principle of associating related keywords with websites that contain these words in search queries. Google's PageRank has revolutionized search by integrating other elements in addition to keyword analysis. How can Google handle large amounts of data based on terabytes and petabytes? The answer is an implementation of computer tools and technologies that are compatible with Hadoop and BigQuery (Google's NoSQL technology).

II. HOW GOOGLE APPLIES BIG DATA TO TRACK AND ANALYZE YOU WITH ITS SERVICES

Google always wanted to develop a search engine that has the ability to think like a human and understand the phrase, logic, and goal of any search query.

Semantics has helped Google to accomplish this task to look beyond the literal meaning of any phrase of a search query. In order to execute the complex process of understanding the user's requirement and preferences, Google has adopted the following techniques-

A. Indexed pages

Indexed pages are the collection of web pages stored to respond to search queries. Indexing is the process of adding web pages into Google search index. It involves assigning keywords or phrases to web pages within a metadata tag or meta-tag so that webpage can be retrieved easily with a search engine that is tailored to search the keywords field. Once the meta-tag is created, Google will crawl and index your webpage. It generally takes 4 days to 4 weeks for any new website to be crawled and indexed by Google.

B. Maintaining the Integrity of the Specifications

Although it doesn't promote itself as such, Google is actually a collection of data and a set of tools for working with it. It has progressed from an index of web pages to a central hub for real-time data feeds on just about anything that can be measured such as weather reports, travel reports, stock market and shares, shopping suggestions, travel suggestions, and several other things.

C. Sorting Tools

Big Data analysis which implies utilizing tools intended to deal with and comprehend this massive data becomes an integral factor whenever users carry out a search query. The Google's algorithms run complex calculations intended to match the questions that user entered with all the available data. It will try to determine whether the user is searching for news, people, facts or statistics, and retrieve the data from the appropriate feed.

D. Knowledge Graph Pages

Google Knowledge Graph is a tool or database which collects all the data and facts about people, places and things along with proper differentiation and relationship between them. It is then later used by Google in solving our queries with useful answers. Google knowledge graph is user- centric and it provides them with useful relevant information quickly and easily.

E. Literal and Semantic Search

The main aim of the literal search engine is to find the root of your search phrase by looking for a match for some of the word or entire phrase. The root of the phrase is then examined and explored upon to display better search results. While semantic search engine tries to understand the context of the phrase by analyzing the terms and language in knowledge graph database to directly answer a question with specific information.

F. Tracking Cookies

Google can keep a track on users across the web by using cookies. If a user is logged or signed into Google and the user is simultaneously browsing other websites, Google can track the websites they are visiting. Google tracks its users across the web by tracking cookies. Thus, Google can collect several data related to users such as their preference, inclination, favorites, requirements etc. Whenever a user searches anything on Google, it incorporates all that information before displaying the results in proper rank.

G. Google Sign-In

The moment you sign in into your Google account, it uses your search history, trends and location to provide accurate search results. Google collects all the data related to the frequency of sites visited, search

phrases used, the timings, data downloaded etc. Google then uses those data to streamline the search results depending upon different scenarios.

H. Google Translate

For complex operations such as translation, Google summons other inbuilt algorithms that are themselves based on Big Data. Google's translate service analyses millions of other pieces of translated text or speech, to determine the most precise interpretation.

I. Google Adwords

Businesses ranging from small scale to large scale are regularly making use of Big Data analytics whenever they advertise through Google Adwords service. Whenever user surfs through different websites, it learns their preferences, likes, dislikes, inclinations etc. on the basis of which Google shows them several advertisements related to products or services that user might be interested in. Advertisers gain admittance to Big Data analytics when they utilize Google Adwords International Journal of Scientific Research in Science, Engineering and Technology (ijsrset.com) 3 and other services such as Google Analytics to lure individuals who fit their customer profile to their sites and stores.

III. BIG DATA SERVICES IN GOOGLE

GCP (Google Cloud Project) offers a wide variety of big data services you can use to manage and analyze your data, including:

A. Google Cloud BigQuery

BigQuery lets you store and query datasets holding massive amounts of data. The service uses a table structure, supports SQL, and integrates seamlessly with all GCP services. You can use BigQuery for both batch processing and streaming. This service is ideal for offline analytics and interactive querying

B. Google Cloud Dataflow

Dataflow offers serverless batch and stream processing. You can create your own management and analysis pipelines, and Dataflow will automatically manage your resources. The service can integrate with GCP services like BigQuery and third-party solutions like Apache Spark.

C. Google Cloud Dataproc

Dataproc lets you integrate your open source stack and streamline your process with automation. This is a fully managed service that can help you query and stream your data, using resources like Apache Hadoop in the GCP cloud. You can integrate Dataproc with other GCP services like Bigtable.

D. Google Cloud Pub/Sub

Pub/Sub is an asynchronous messaging service that manages the communication between different applications. Pub/Sub is typically used for stream analytics pipelines. You can integrate Pub/Sub with systems on or off GCP, and perform general event data ingestion and actions related to distribution patterns.

E. Google Cloud Composer

Composer is a fully-managed cloud-based workflow orchestration service based on Apache Airflow. You can use Composer to manage data processing across several platforms and create your own hybrid environment. Composer lets you define the process using Python. The service then automates processing jobs, like ETL.

F. Google Cloud Data Fusion

Data Fusion is a fully-managed data integration service that enables stakeholders of various skill levels toprepare, transfer, and transform data. Data Fusion lets you create code-free ETL/ELT data pipelines using a point-and-click visual interface. Data Fusion is an open source project that provides the portability needed to work with hybrid and multicloud integrations.

G. Google Cloud Bigtable

Bigtable is a fully-managed NoSQL database service built to provide high performance for big data workloads. Bigtable runs on a low-latency storage stack, supports the open-source HBase API, and is available globally. The service is ideal for time-series, financial, marketing, graph data, and IoT. It powers core Google services, including Analytics, Search, Gmail, and Maps.

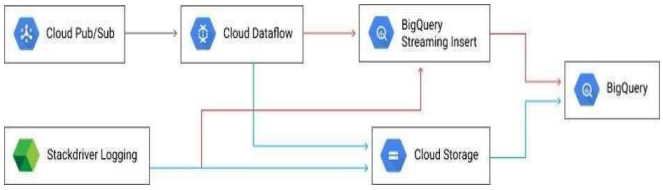


Figure 1 - Big Query Storage [3]

IV. WHAT IS BIG QUERY?

Google's BigQuery is an enterprise-grade cloud-native data warehouse. BigQuery was first launched as a service in 2010 with general availability in November 2011. Since inception, BigQuery has evolved into a more economical and fully-managed data warehouse which can run blazing fast interactive and ad-hoc queries on datasets of petabyte-scale. In addition, BigQuery now integrates with a variety of Google Cloud Platform (GCP) services and third-party tools which makes it more useful.

Level architecture: It has a columnar storage and tree architecture of Dremel, BigQuery offering unprecedented performance. As illustrated below, a BigQuery client (typically BigQuery Web UI or bg command-line tool or REST APIs) interacts with Dremel engine via a client interface.

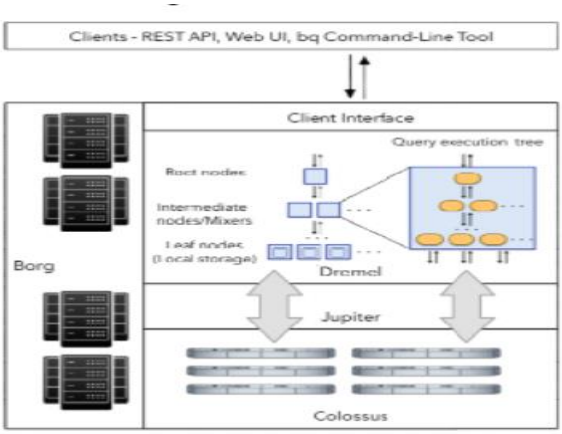


Figure 2 - Client Interface [5]

A. Storage:

You can import your data into storage via Batch loads or Streaming. During the import process, BigQuery encodes every column separately into Capacitor format. Once all column data is encoded, it's written back to Colossus. During encoding various about the data is collected which is later used planning.

B. Execution Model:

In a serving tree, a root server receives incoming queries from clients and routes the queries to the next level. The root server is return query results to the client. Leaf nodes of the tree do the heavy lifting of reading the data from and performing filters and partial aggregation.

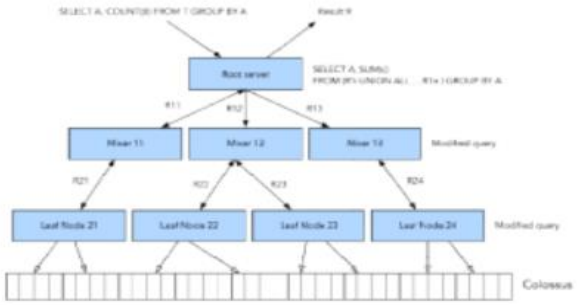


Figure 3 – Leaf Node Execution [5]

Each leaf node provides execution thread or number of processing units often called as slots. It is important to note, BigQuery automatically calculates how many slots should be assigned to each query. The number of allocated slots are assigned depending on query size and complexity.

V. CONCLUSION

Google isn't just influencing the way it is today. They analyze big datasets (MapReduce, BigQuery, etc.), but they are probably always responsible for making it part of us more than anyone else. Most companies doing it today will do it for years to come. Google collects vast amounts of data about the people who use it. It uses AdSense algorithms to match your online business with potential customers. Companies are willing to pay for these recommendations that appear as ads in their customers' browsers.

Another big data project that Google is working on is the self-driving car. It uses and produ es large amounts of data from the combination of sensor with cameras, analysis from Google Maps. Street View and other sources allow Google cars to drive safely on the road without input from human drivers. Google may not be ready to predict the future, but the position of as a key player and innovator for big data seems safe.

VI.REFERENCES

[1]. https://www.analyticssteps.com/blogs/companies data
[2]. https://medium.com/xnewdata/Google success-case- 949e1b99e678
[3]. https://tweakyourbiz.com/technology/
[4]. https://www.techtarget.com/searchcloudcomputing/featur e/Wor ds-to-go-Google-big-data-analytics
[5]. Jun, Seung-Pyo, Hyoung Sun Yoo, and San Choi. "Ten years of research change using Google perspective of big data utilizations and applications." Technological forecasting and social change 130 (2018): 69-87.
[6]. De Falco, Ciro Clemente, Noemi Crescentini, and Marco Ferracci. "The Spatial Dimension in Social Media Analysis."

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Study on Robotics and Nano Technology in the Field of Medical Science

Dr Senthil Kumar R[1], Balaji S[2], Ammidal Rakesh[2], Amidala Vinay Ranga[2], B N Shivacharan[2]

[1]Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]UG Student, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

One of the 21st century's most promising technologies is nanotechnology. Nanomedicine is a nanotechnology, refers to highly specific medical intervention at the molecularsscale for curing disease or repairing damaged-tissues, such as bone, muscle, or nerve. Nanotechnology is a collective term referring to technological developments on the nanometer scale, usually 0.1 to 100 nm. A nanometer is one-billionth of a meter, too small to be seen with a conventional laboratory microscope. It is at this size scale - about 1000nanometers or less - that biologicalmmolecules and structures inside living cells operate. Therefore, nanotechnology is engineering and manufacturing at the molecular scale. Utilities of nanotechnology to biomedicalssciences which is designed to interact with the body at sub-cellular scales with a high degree of specificity. The idea is potentially translated into targeted cellular-tissue-specific clinical applications aimed at maximal therapeutic effects with very limited adverse-effects. The paper shows an overview of some of the applications of nanotechnology in nanomedicine.

Keywords— Nano medicine, Nano technology

I. INTRODUCTION

Nanomedicine offers the prospect of powerful new tools for the treatment of human diseases and the improvement of human biological systems using molecular Nanotechnology. The term 'Nanotechnology' generally refers to engineering and manufacturing at the molecular or nanometer length scale (A nanometer is one-billionth of a meter, about the width of 6 bonded carbon atoms).

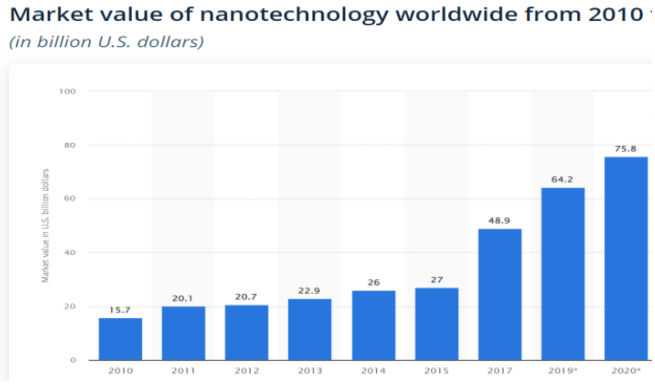


Fig. 1 Market value of nanotechnology

Nanotechnology, 'the manufacturing technology of the 21st century', will provide an opportunity to build

a broad range of economically complex molecular machines (including, not incidentally, molecular computers). It will lead to the building of computer controlled molecular tools much smaller than a human cell with the accuracy and precision of drug molecules. Such tools will allow medicine, for the first time to intervene in a sophisticated and controlled way at the cellular and molecular level.

A growing interest in the medical applications of nanotechnology has led to the emergence of a new field called nanomedicine. They could remove obstructions in the circulatory system, kill cancer cells, or take over the function of subcellular organelles. Just as today the artificial heart has been developed, so in the future, perhaps artificial mitochondrion would be developed.

Nanomedicine offers the prospect of powerful new tools for the treatment of human diseases and the augmentation of human biological systems. Diamondoid-based medical nanorobotics may offer substantial improvements incapabilities over natural biological systems, exceeding even the improvements possible via tissue engineering and biotechnology. This review is aimed to explore several of the worst medical problems and how nanotechnology can be used to cure them.

Diseases and its Cures by Nanomedicine

Medical science has scored some impressive successes. Antibiotics have reduced diseases caused by bacteria remarkably. Nowadays, vitamin and mineral deficiency diseases are rare in developed nations.

However, there are still many diseases that limit our lifespan, and the medicines concerned can only postpone them but are not able to cure. Life cannot be extended indefinitely without curing each disease that threatens to shorten it.

Treatment of Cancer

At a cellular level, cancerous tissues are usually quite different from normal tissues. Many cancer cells change the chemicals on their surface, and are therefore easy to identify. However, most cancer cells grow faster or change shape and every cancer involves a genetic change that causes a difference in the chemicals inside the cell.

The immune system takes advantage of surface markers to destroy cancer cells; but this is not enough to keep us cancer-free. Nanobots will have several advantages. Firstly, they can physically enter cells and scan the chemicals inside.

Table 1 A Partial Nanomedicine Technologies Taxonomy

Raw nanomaterials	**Cell simulations and cell diagnostics**	**Biological research**
Nanoparticle coatings	Cell chips,	Nanobiology
Nanocrystalline materials	Cell stimulator	Nanoscience in life sciences
Nanostructured materials	**DNA manipulation, sequencing, diagnostics**	**Drug delivery**
Cyclic peptides	Genetic testing	Drug discovery
Dendrimers	DNA microarrays	Biopharmaceutics
Detoxification agents	Ultrafast DNA sequencing	Drug encapsulation
Drug encapsulation	DNA manipulation and control	Smart drugs
Fullerenes	**Tools and diagnostics**	**Molecular medicine**
Functional drug carriers Smart drugs	Bacterial detection systems	Genetic therapy
MRI scanning (nanoparticles)	Biochips	Pharmacogenomics
Nanobarcodes	Biomolecular imaging	**Artificial enzymes and enzyme control**
Molecular medicine	Biosensors and biodetection	Enzyme manipulation and control
Nanoemulsions	**Diagnostic and defense applications**	**Nanotherapeutics**
Nanofibers	Endoscopic robots and microscopes	Antibacterial and antiviral nanoparticles
Nanoparticles	Fullerene-based sensors	Fullerene-based pharmaceuticals
Nanoshells	Imaging (cellular, etc.)	Photodynamic therapy
Carbon nanotubes	Monitoring	Radiopharmaceutica ls
Noncarbon nanotubes	Lab on achip	**Synthetic biology and early nanodevices**
Quantum dots	Nanosensors	Dynamic nanoplatform
Artificial binding sites	Point of care diagnostics	nanosome
Artificial antibodies	Protein microarrays	Tecto-dendrimers
Artificial enzymes	Scanning probe microscopy	Artificial cells and liposomes
Artificial receptors	**Intracellular devices**	Polymeric micelles and polymersomes
Molecularly imprinted polymers	Intracellular biocomputers	**Biotechnology and biorobotics**
Control of surfaces	Intracellular sensors/reporters	Biologic viral therapy
Artificial surfaces adhesives	Implants inside cells	Virus-based hybrids
Amficial surfaces— nonadhesive	**BioMEMS**	Stem cells and cloning
Artificial surfaces— regulated	Implantable materials and devices	Tissue engineering
Biocompatible surfaces	Implanted bioMEMS, chips, and electrodes	Artificial organs
Biofilm suppression	MEMS/Nanomaterial s-based prosthetics	Nanobiotechnology
Engineered surfaces	Sensory aids (artificial retina, etc.)	Biorobotics and biobots
Patterned surfaces.	Microarrays	**Nanorobotics**
	Microcantilever-based sensors	DNA-based devices and nanorobots
	Microfluidics	Diamond-based nanorobots
	Microneedles	Cell repair devices
	Medical MEMS	
	MEMS surgical devices.	

Secondly, they can have on board computers that allow them to do calculations not available to immune cells. Thirdly, nanobots can be programmed and deployed after a cancer is diagnosed, whereas the immune system is always guessing about whether a cancer exists.

Nanobots can scan each of the body's cells for cancerous tendencies, and subject any suspicious cells

to careful analysis; if a cancer is detected, they can wipe it out quickly, using more focused and vigorous tactics than the immune system is designed for.

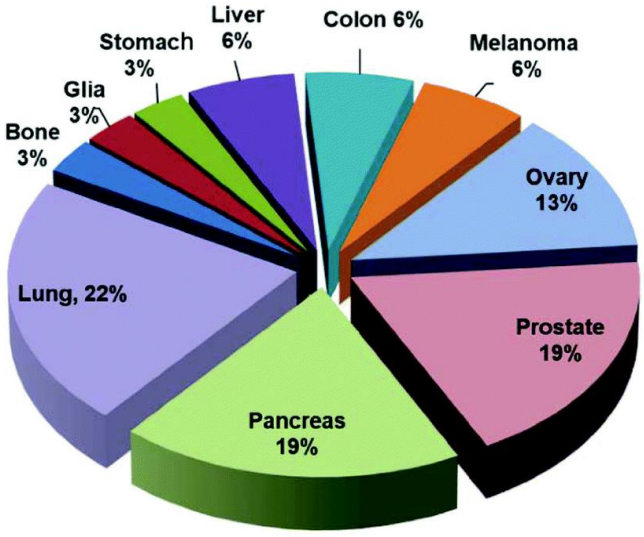
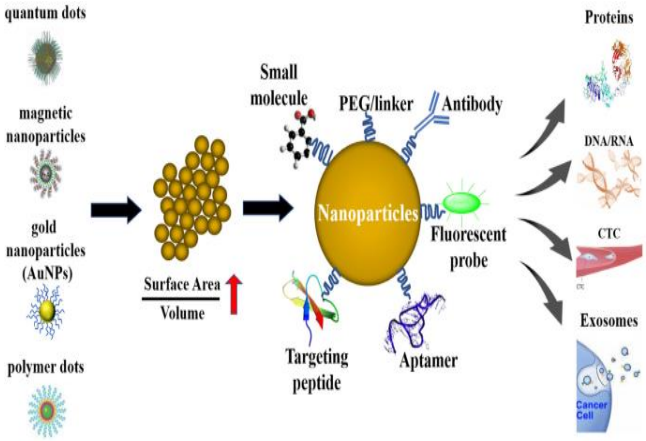


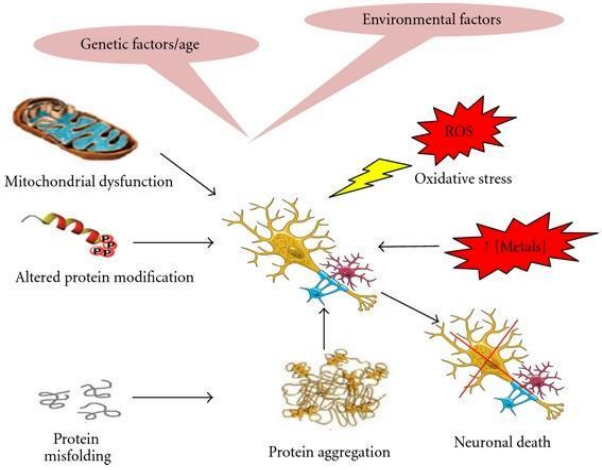
Fig. 2: Effectiveness Nano technology for treating various types of cancers.



Prevention of Brain Damage in Neurodegenerative Diseases

The brain is unique among the body's organs; it stores our memories and personality, and therefore, t it cannot simply be replaced if it starts to wear out. This poses a special problem for life extension; the information stored in the brain must be preserved over extended periods of time, safe from disease and accident. Obviously, it is good to prevent the premature death of neurons. Poisons such as alcohol, accidents such as stroke, and diseases such as Alzheimer's can all cause neurons to die. In each of these cases, neuron death can be greatly slowed if not prevented entirely by controlling the chemistry inside the cell. Injurious chemicals can be vacuumed up and converted into harmless ones.

Damaged neurons, like other cells, sometimes go into suicide mode (called "apoptosis"); as mentioned above, this can be chemically prevented, and the neuron can be stabilized until the problem is fixed and the damage is repaired. It is now acknowledged that brain cells do regenerate; the brain is generating new cells all the time. This implies that some neural death is normal. How do the new cells know how to behave? It seems that a new neuron can take its cues from the existing ones; this means that a person's mind may be intact even after the death and replacement of a large percentage of their neurons. Finally, it may be possible to measure neural connections and/or activity in enough detail to simulate the firing pattern. This may make it possible to create an artificial neuron or even an artificial neural net that can be used to replace missing neurons and retain old memories. But even if this proves to be impossible, the worst-case scenario is one in which people cannot remember much farther than a century back. More memory loss than this can be accepted as a natural consequence of aging.



Hormone Deficiency

Aging is associated with changes in the levels of many hormones; perhaps the best known example is menopause, which is caused by a reduction in estrogen. It is likely that treating glands against aging at the cellular level would restore age-appropriate hormone production. However, if this is not enough to bring the body to a younger state, artificial glands could be built that would maintain the desired hormone levels. In fact, different hormone levels could be supplied to different organs -something that the body cannot do for itself. This would be an example of heterostasis.

For the treatment of Infection

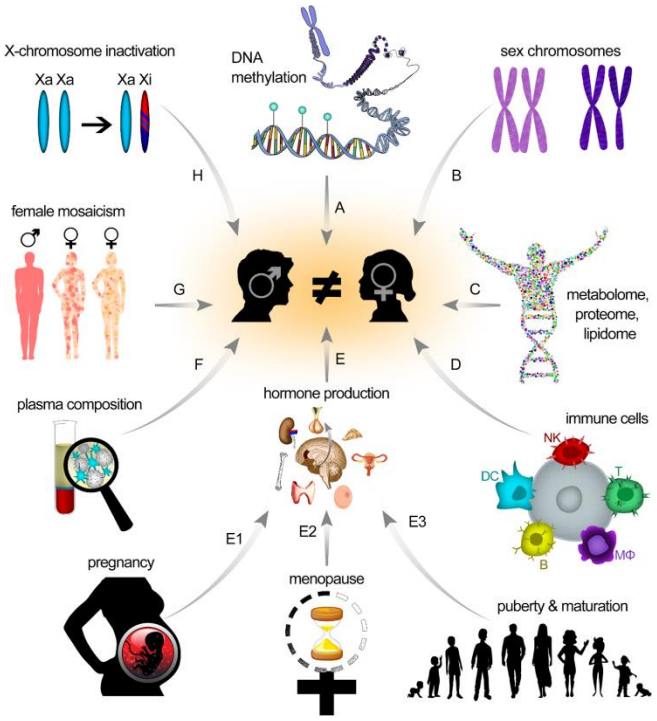
Bacteria, viruses, and parasites are continuing problems. Antibiotics work well against most bacteria; however, antibiotic-resistant strains are developing. Since viruses are not active until they take over a cell, they are immune to antibiotics, and medicine cannot do much against them. There are many kinds of parasites that may need individual medical techniques. Our immune system is quite effective in dealing with most infections but the immune system needs to learn about it by experience.

It is generally most effective at fighting organisms that it can recognize on a molecular level. Some diseases, such as Ebola, progress too rapidly for the immune system to respond. Syphilis survives by being stealthy and by surrounding itself with the chemicals of the body to camouflage itself. Herpes splices itself into the genes of the body's cells, so the immune system cannot detect it and wipe it out. HIV directly attacks the immune system. Nanobots have several advantages over the immune system. They will not be susceptible to attack by natural pathogens. They will have computational resources unavailable to immune cells. They can be programmed to find and fight diseases they have never encountered (when a new disease shows up), as soon as it is analyzed everyone's nanobots can benefit.

Likewise, the system can be activated based on external knowledge of the likelihood of a disease; the nanobots would not have to waste energy looking for malaria in winter. Nanotech will provide more options for cleaning up after a disease, since corrupted genes will be repairable without killing the affected cell. Some diseases, such as cholera and tetanus, live in the environment; without scrubbing the whole earth, we cannot get rid of them entirely, therefore, we will need to maintain an immune system against them. But many diseases cannot survive without humans to infect.

With great effort, we managed to eradicate smallpox using 1970's technology. Cheap manufacturing would allow the creation of billions of doses of highly effective treatments that would be easy to distribute and administer; the main obstacles to wiping out many diseases worldwide would be political, not economic or technological. Nanorobotic 'Microbivores' the nanorobotic phagocytes (artificial white cells) traveling in the bloodstream could be 1000 times faster-acting than white blood cells and eradicate 1000 times more bacteria, offering a complete antimicrobial therapy without increasing the risk of sepsis or septic shock (as in traditional antibiotic regimens) and without the release of biologically active effluents. Microbivores could also be useful for treating infections of the meninges or the cerebrospinal fluid (CSF) and respiratory diseases involving the presence of bacteria in the lungs or sputum, and could also digest bacterial biofilms. These handy nanorobots could quickly rid the blood of nonbacterial pathogens such as viruses (viremia), fungus cells (fungemia), or parasites (parasitemia). Outside the body, microbivore derivatives could help clean up biohazards, toxic biochemicals or other environmental organic materials spills, as in

bioremediation



For Life Saving after Accidents

Accidents, especially motor vehicle accidents, are a leading cause of death in all ages. Although an accident is not itself a disease, it kills as it damages the body, and that damage can be treated or prevented like a disease. Most accidents involve mechanical injury (trauma) and many involve chemical injury, either by poisoning or oxygen starvation.

A permanent nanorobot installation can ensure survival. Nanobots embedded in a tissue can strengthen it against tearing, or repair once it is torn. It is common for a blow to the head to rattle the brain against the skull; a specially shaped nano-built device could cushion the brain, preventing this damage. Other devices could vacuum up common poisons before they could cause damage, or barricade poisoned areas to keep the poison from spreading through the body. 'Respirocytes' the artificial red blood cells could allow the body to function normally for several minutes without breathing or blood circulation, giving more opportunity to restore normal functioning. In cases of extreme injury, heterostasis could be used to stabilize the body until help can arrive. As long as the brain is not physically damaged, it can be functionally separated from the body and forced into a low-power state.

With today's medicine, paramedics refer to the "golden hour"; if an accident victim can be brought to a hospital in less than an hour,the chance of survival is greatly increased. People have recovered after drowning in cold water for over an hour; artificial mimicry of this state, combined with the ability to aggressively repair the body, might extend the "golden hour" significantly. 'Respirocytes' - artificial red blood cells comprised of microscopic diamondoid pressure tanks that are operated at up to 1000 atmospheres of pressure - could carry greater than 200 times more respiratory gases than an equal volume of natural red blood cells. The injection of a 5-mililitre therapeutic dose of 50% respirocyte saline suspension, a total of 5 trillion individual nanorobots, into the human bloodstream would exactly duplicate the gas-carrying capacity of the patient's entire 5.4 liters of blood.

Primary medical applications of respirocytes would include transfusable blood substitution; partial treatment for anemia, perinatal/neonatal, and lung disorders; enhancement of cardiovascular/neurovascular procedures, tumor therapies and diagnostics; prevention of asphyxia; artificial breathing; and a variety of sports, veterinary, battlefield, and other uses.

Fig 3: Tissue Nanotransfection

For Blood-Related Diseases

Many diseases, from heart attacks and strokes to sepsis and metastasizing cancer, involve the blood in some way. The author has proposed an aggressive nanomedical device, a "Vasculoid" that would replace the blood volume and take over its functions by lining the entire vascular system with a multi-segmented robot. In addition to preventing many diseases, and limiting the scope of others (such as poisoning), such a system would provide detailed control of the body's chemical environment around each individual capillary, allowing heterostasis to be used extensively.

A second application would be to provide metabolic support in the event of impaired circulation. Poor blood flow, caused by a variety of conditions, can result in serious tissue damage. A major cause of tissue damage is inadequate oxygen. A simple method of improving the levels of available oxygen despite reduced blood flow would be to provide an "artificial red blood cell." We will consider a simple design here: a sphere with an internal diameter of 0.1 microns (100 nanometers) filled with high pressure oxygen at ~1,000 atmospheres (about 108 pascals). The oxygen would be allowed to trickle out from the sphere at a constant rate (without feedback). Diamond has a Youngs modulus of about 1012 pascals. An atomically precise diamondoid structure should be able to tolerate a stress of greater than 5 x 1010 pascals (5% of the modulus). Thus, a 0.1 micron sphere of oxygen at a pressure of 108 pascals could be contained by a hollow diamondoid sphere with an internal diameter of 0.1 microns and a thickness of less than one nanometer.

While providing oxygen to healthy tissue should maintain metabolism, tissues already suffering from ischemic injury (tissue injury caused by loss of blood flow) may no longer be able to properly metabolize oxygen. In particular, the mitochondria will, at some point, fail. Increased oxygen levels in the presence of nonfunctional or partially functional mitochondria will be ineffective in restoring the tissue. However, more direct metabolic support could be provided. The direct release of ATP, coupled with selective release or absorption of critical metabolites (using the kind of selective transport system mentioned earlier), should be effective in restoring cellular function even when mitochondrial function had been compromised. The devices restoring metabolite levels, injected into the body, should be able to operate autonomously for many hours (depending on power requirements, the storage capacity of the device and the release and uptake rates required to maintain metabolite levels). Advances in medical technology necessarily depend on the understanding of living systems. With the kind of devices discussed earlier, it would be possible to explore and analyze living systems in greater detail than ever considered possible before.

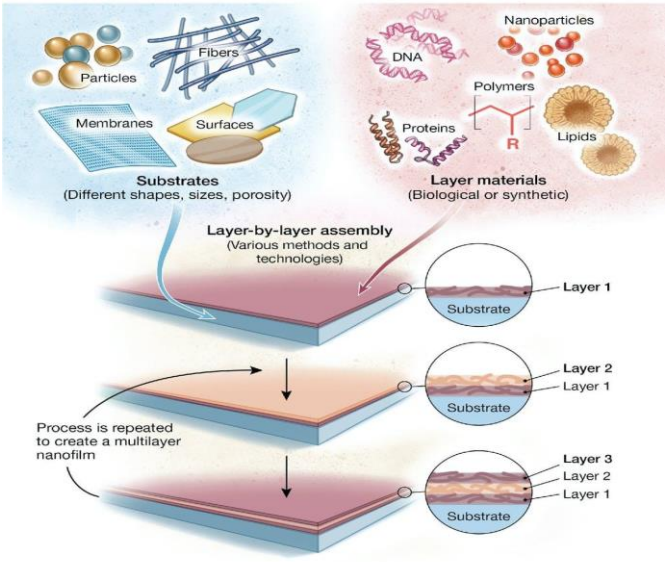
Autonomous molecular machines, operating in the human body, could monitor levels of different compounds and store that information in internal memory. They could determine both their location and the time. Thus, information could be gathered about changing conditions inside the body, and that information could be tied to both the location and the time of collection. Physical samples of small volumes (nano tissue samples) could likewise be taken. These molecular machines could then be filtered out of the blood supply and the stored information (and samples) could be analyzed. This would provide a picture of activities within healthy or injured tissue. This new knowledge would give us new insights and new approaches to curing the sick and healing the injured.

In the first half of the 21st century, nanomedicine should eliminate virtually all common diseases of the 20th century, and virtually all medical pain and suffering as well. Only conditions that involve a permanent loss of personality and memory information in the brain – such as an advanced case of

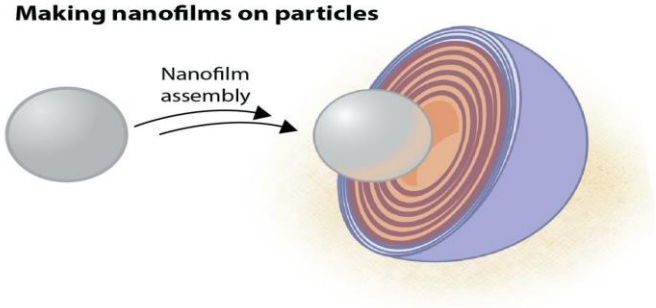
Alzheimer's disease or a massive head trauma may remain incurable in the nanomedical era. Because aging is believed to be the result of a number of interrelated molecular processes and malfunctions in cells, and because cellular malfunctions will be largely reversible, middle-aged and older people who gain access to an advanced nanomedicine can expect to have most of their youthful health and beauty restored. Hence they may find few remaining limits to human

In the longer term, perhaps 10–20 years from today, the earliest molecular machine systems and nanorobots may join the medical armamentarium, finally giving physicians the most potent tools imaginable to conquer human disease, ill-health, and aging. The future prospect of nanomedicine is bright and we are required to work very hard and long to make it fruitful.

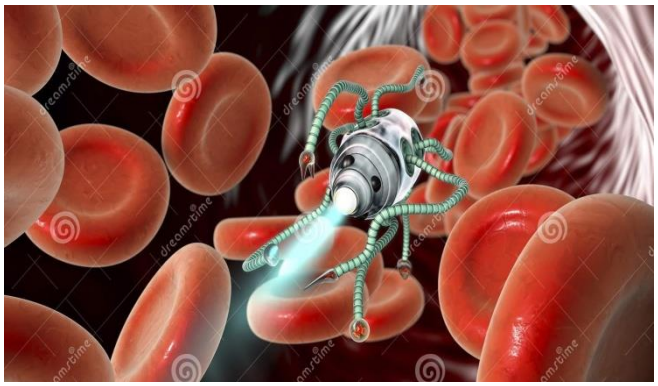
Making nanofilms



Making nanofilms on particles



Nano Bots in blood stream



II. REFERENCES

[1]. FreitasRA.-Jr The futureoof molecular scale devicesiin nanomedicine. Stud-Technol-Inform 2002-80-45-59 [PubMed]

[2]. Ralph C. Merkle. Nanotechnology and Medicine, Advances in Anti-Aging Medicine, Vol. I, edited by Dr. Ronald M. Klatz, Liebert press, 1996. p. 277-286.

[3]. Freitas RA Jr Nanomedicine,VVolume I: Basic Capabilities, LandesBioscience,Georgetown, TX,1999;

[4]. Freitas RA Jr. The Nanomedicine Page, http://www.foresight.org/Nanomedicine/index.html

[5]. Freitas RA., Jr What is nanomedicine? Nanomedicine: NBM 2005;1:2-9 [Google Scholar]

[6]. Bharali-DJ, Khalil-M, GurbuzMM, Simone TM, Mousa SA. Nanoparticles and cancer therapy: a conciserreview with emphasis onddendrimers. Int J Nanomedicine 2009;4:1-7 [PMC free article] [PubMed] [Google Scholar]

[7]. Park JW. Liposome-based drug delivery in breast cancer treatment. Breast Cancer Res 2002;4(3):95-99 10.1186/bcr432 [PMC-article] [PubMed] [CrossRef] [Google Scholar]

[8]. Thorek DL, ChenAK, Czupryna J, TsourkasA.Superparamagnetic iron oxide-nanoparticle probes for molecular imaging. Ann

Biomed-Eng 2006. Jan;34(1):23-38 10.1007/s10439-005-9002-7

[9]. Partha R, Conyers JL. Biomedical applications of functionalized fullerene-based nanomaterials. Int J Nanomedicine 2009;4:261-275 [PMC-free-source article].

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Social Distance Detection Using CNN

Nikitha P, Krithin R, P M ArshiyaZeba, Talamadupula Ameesha
UG Student, Department of Computer Science and Engineering, New Horizon College of Engineering, Marathalli, Bengaluru, Karnataka, India

ABSTRACT

Coronavirus is a dangerous pandemic that has thrown the entire planet into chaos. As of October 23, 2020, more than 42 million people have been poisoned, with 1.14 million deaths reported worldwide. A deeper understanding of the pandemic suggests that a single person's carelessness can have far-reaching consequences that are difficult to argue. As is often said, prevention is preferable to cure, and the WHO has advocated a few security measures to decrease COVID spread. In the current situation, social separation has shown to be one of the most effective spread plug techniques. Social separating is also known as "physical separating," which involves keeping a distance between yourself and those around you. The use of social removal aids in the reduction of actual contact or collaboration between COVID-19-affected people and healthy people. To follow the social separating, everyone should keep a distance of about 6 feet between them, according to WHO's standard remedies. This is an obvious way to break the viral chain. As a result, each of the effected countries has seen societal separation.

I. INTRODUCTION

Observing social distance in circumstances over time is a challenging task. It is possible to do so in two ways: automatically and manually. The manual technique necessitates a large number of genuine eyes to ensure that everyone is strictly adhering to social distancing rules. This is a challenging interaction because one cannot save their eyes for constant checking 24 hours a day, seven days a week. CCTV cameras have replaced many human eyes with automated or autonomous observation systems. CCTV cameras record video, which is examined by a robotized observation framework. When any suspicious event occurs, the framework raises an alarm. Security personnel can make important movements with this information at their disposal. As a result, the automated checking framework has outperformed a number of the manual observing technique's limitations.

The goal of this investigation is to limit the impact of the COVID epidemic to financial old rarities with minimal harm. We have proposed a compelling programmed observation framework in this paper that aids in the identification of each individual and screens them for the social distancing boundary. This application works well in both indoor and outdoor observation settings. It is commonly found in a variety of locations, including train stations, airports, mega stores, shopping malls, and roadways[1]. The proposed method can be thought of as a combination of two basic errands:

(i) Human detection and tracking
(ii) Monitoring of social distancing among humans

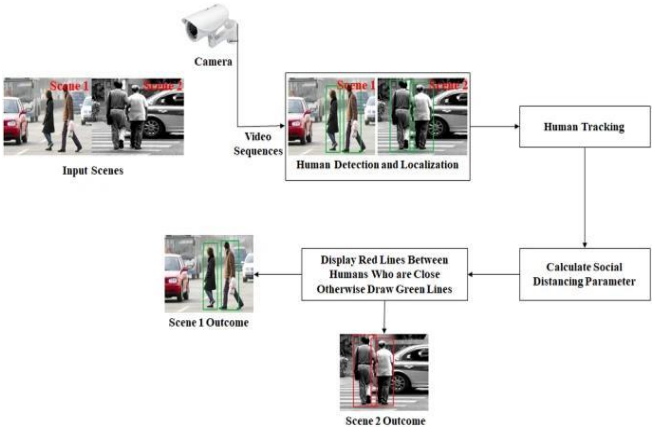


Fig. 1. Simple recognition workflow

The overall scenario of publicly observing for social separating is introduced, with CCTV cameras available at open spots for reconnaissance and social separating observation. The item identification and following module takes care of the video transfer/outline successions obtained from these cameras in order to find human presence in the scene. For evaluating the level of social separating rehearsed, boundaries such as the 'centroid' of individual area and the 'distance' among many such centroids are analysed. The colour of the bouncing box of people recognised changes from green to red, signalling an alarm. Until there is a suitable space between two persons, the bouncing box's shade is green[2]. As this decreases, the colour of the jumping boxes turns red, indicating a breach of social separation. The proposed method could still be improved to ensure that the numbers are linked and to alert us when distance requirements are not observed. To make the workflow more efficient, the following Python libraries are used: - • YOLO Transfer Learning Process • NumPy • OpenCV2 • SciPy

Different Build Approach

This technique for detecting social distancing was created to identify the safety distance between people in public areas. In this study, the deep CNN approach and computer vision techniques are used. Initially, the pedestrian in the video frame was detected using an open-source object detection network based on the YOLOv3 method[9]. Only the pedestrian class was used as a result of the detection, and other object types were ignored in this application. As a result, the bounding box that best fits each identified pedestrian can be drawn in the image, and this data will be utilised to calculate distance.When it comes to deep learning-based object identification, the YOLO model is one of the state-of-the-art models that has been shown to deliver considerable speed improvements and is suited for real-time applications. The YOLO model was used to detect pedestrians in this study.

NumPy is Python's core package for logical registering. It's a Python package that provides a multidimensional exhibit object, several inferred objects (such as veiled clusters and grids), and a set of schedules for rapid cluster operations, including numerical, graphical, and textual. I/O, discrete Fourier changes, necessary straight, sensible, shape control, organising, choosing arbitrary reproduction, variable-based math, fundamental measurable activities. The ndarray object is at the heart of the NumPy bundle. This exemplifies n-dimensional variants of homogeneous information types, with a wide range of operations enacted in assembled code for execution.

There are a few key differences between NumPy exhibitions and ordinary Python displays: Unlike Python records, NumPy exhibits have a fixed size when they are created (which can develop powerfully). When you change the size of anndarray, it creates a new exhibit and deletes the old one. The components of a NumPy display must all be of the same information type and, as a result, be of the same memory size. The particular case: several types of (Python, including NumPy) objects can be created,

taking into consideration different types of estimated components. NumPy shows how to work with cutting-edge numerical and other types of procedures on massive amounts of data. In most cases, similar operations can be completed more quickly and with less code than using Python's underlying groupings.

Argphrase - As a replacement for optparse, the argparse module was added to Python 2.7. The execution of argparse maintains highlights that would not have been difficult to add to optparse, and that would have necessitated in reverse inconsistent Programming interface changes, therefore, all things being equal, another module was added to the library. optparse is still supported, although it is unlikely to acquire any new features. The programming interface for argparse is similar to that of optparse, and argparse may often be used as a direct replacement by simply changing the names of the classes and strategies used. If argparse is available on all arrangement stages, new projects should most likely use it.

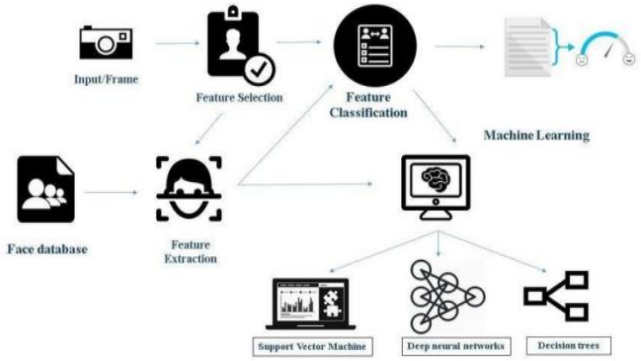


Fig. 2. Applications of OpenCV .

OpenCV is a cross-stage library utilizing which we can foster constant PC vision applications. It primarily centers around picture handling, video catch and examination including highlights like face discovery and article location[6][7]. Utilizing OpenCV library, you can – Peruse and compose pictures Catch and save recordings Interaction pictures Perform include discovery Identify explicit articles like faces, eyes, vehicles, in the recordings or pictures. Break down the video, i.e., gauge the movement in it, take away the foundation, and track objects in it.

How the Model Serves

It delays the peak of the epidemic and gives health professionals more time to prevent flu transmission. It also allows time to organize important clinical conveniences such as a brow thermometer, an infrared thermometer, an FFP2 veil, an FFP3 veil, a Face veil, a Careful Face veil, a KN95, a N95, a Defensive Cover, mattresses, ventilators, and so on. It reduces the number of contaminants in people. It disperses the quantity of enticing persons over a longer period of time. Separating people socially slows the spread of illness and can even bring an episode to a halt. When sickness is spread through physical contact, such as hacking or wheezing, this is the recommended course of action.But in some cases, when disease is spread mostly by contaminated water or food, mosquitoes, or creepy crawlies, it has a lower success rate. If people are modified in accordance with the human communication atmosphere, it might produce despondency and regular depression. It reduces the utility of human communication and results in the loss of a variety of benefits[8].

python is effective

Python is a widely used, high-level programming language that is extremely useful. It was designed in 1991 by Guido van Rossum and afterwards by the Python Programming Establishment. It was designed with code coherence in mind, and its sentence structure allows software engineers to explain their ideas in fewer lines of code. Python is a programming language that enables you to work quickly and more effectively organize frameworks.Python may be used to create web apps by a specialist. Python can be used in conjunction with programming to create work measurements. Python has the ability to

communicate with database systems. It can also inspect and modify records. Python may be used to manage large amounts of data as well as execute complex calculations. Python can be used for rapid prototyping as well as creation-oriented programming.Python eats away at the problem in phases (Windows, Macintosh, Linux, Raspberry Pi, and so on). Python follows the same core structure as English. Python is based on a translation framework, which states that code can be executed once it has been assembled. This demonstrates that prototyping may be done in a very short amount of time. Python can be approached in a procedural, article-based, or utilitarian fashion.

Python is a straightforward programming language. It's simple to read, compose, study, and grasp.Python is known for its ease of learning and adaptability. Learning is not difficult. Python has a simple punctuation system, and Python programming is also simple. Because it's so simple, you can easily read and comprehend other people's code.

Reading Python code is similar to reading an English sentence. This is one of the main reasons why Python is the best language for beginners. In contrast to other programming languages, Python uses space rather than wavy supports. This improves the code's appearance and clarity. Python is a language that has been decoded. It is used in conjunction with the Inactive (Intelligent Advancement Climate). This is a translator that adheres to the REPL format (Read-Assess Print-Circle)[14]. It runs each line and displays the yield for each one. As a result, it detects errors while you're executing a line and displays the entire stack as a result of the error. In the event that you require assistance, the local area will be able to assist you. They've also already answered a lot of Python-related questions.

Python is open-source, and the community is always adding to it in order to improve it. It is open source and its source code is freely available to the general public. Python can be downloaded from the official Python website. The standard library is massive, with a plethora of bundles and modules that are both common and useful. If you require something that is available in this standard library, you do not need to compose it without first preparing it. Along these lines, you can concentrate on more important issues. C++ is a mid-level language, while Python is a high level language. It's simple and gets you closer to the customer. You don't have to worry about remembering framework engineering or dealing with memory. The Python Standard Library is a collection of Python's accurate grammar, tokens, and semantics. It comes with a Python transporter in the middle. We used this as a starting point for a presentation. It's written in C and takes care of things like I/O and other core modules. Python is what it is because of all of these features. The core of the standard library consists of about 200 centre modules.

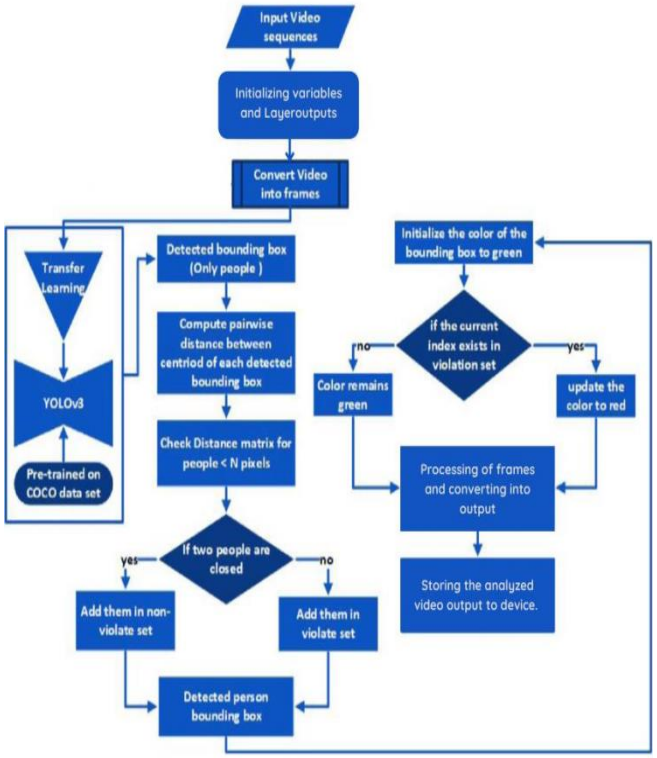


Fig. 3. Modeled architecture

II. DIGITAL IMAGE PROCESSING

This is the first step or interaction of advanced picture preparation's basic steps. Getting a picture that is now in electronic format and safeguarding it might be as simple as getting an image that is now in computerized format. Preprocessing, such as scaling and so on, is usually included in the photo getting step. Picture enhancement is one of the most straightforward and enjoyable aspects of sophisticated image preparation. Essentially, the idea behind upgrade methods is to bring out detail that has been obscured, or to highlight certain areas of importance in a photograph. For instance, altering brilliance and distinction, and so forth. Picture restoration is a field that also deals with restoring a picture's appearance[3]. In any case, unlike upgrade, which is subjective, picture reclamation is objective, as rebuilding procedures would typically be based on numerical or probabilistic models of image degradation. Due to the massive increase in the use of computerized images over the Internet, the field of shading picture preparation has grown in importance. This could include shading displaying and preparation in a computerized environment, for example. Wavelets are the foundation for addressing images at various levels of goal. For information pressure and pyramidal depiction, pictures are divided into smaller districts. Pressure manages tactics for reducing the amount of storage space or data transfer required to send a picture. It is quite crucial to bundle information, especially in web-based jobs.

Morphological preprocessing deals with tools for extracting image segments that are useful in the depiction and representation of shape. A image is divided into its essential elements or items by division systems.[5] Self-governing division is, by and large, one of the most difficult tasks in sophisticated image production. A difficult division system takes the interaction a long way toward a fruitful arrangement of imaging issues that require things to be recognized on their own. Extraction of ascribes that result in some quantitative data of interest or are essential for differentiating one class of objects from another is managed by depiction.

Information could be as simple as listing locations on an image where the data of interest is known to be found, hence limiting the amount of time spent seeking for that data. Furthermore, the information base can be quite unreliable[15]. For example, in a materials review issue, an interconnected rundown of all significant probable deformities, or a picture data set providing high-goal satellite photos of a district for change-location applications.

USE OF Convolutional Neural Networks

Artificial Intelligence (AI) includes subdivisions such as machine learning (ML) and deep learning (DL). The following innovation in Machine Learning is addressed by Profound Learning. In Deep Learning, the model learns by simulating a human cerebrum's neuronal system. This allows the model to study information in a structure similar to how people do. A human developer isn't required to mediate and decide how to manage information in Profound Learning models[4]. It is well-equipped to benefit from the unusual amount of data sent to it. PC vision allows a computer to see in the same way that humans do. It's a type of software engineering that focuses on reproducing the inconsistencies of the human visual framework. It uses a computer to recognize and measure the items in photos. Superhuman precision in picture arrangement, object location, picture reclamation, and picture division has been achieved with deep learning. It employs massive neural networks to teach machines how to automate the tasks performed by human visual systems[11]. It is a

field that entails gaining a thorough understanding through the use of enhanced images or recordings.

C Vision can be defined as a discipline that explains how to reconstruct, interfere with, and comprehend a 3D scene from its 2D images in terms of the design qualities contained in the scene. It uses computer programming and equipment to display and duplicate human vision. PC Vision encompasses the following fields in its entirety :

Picture Control is at the heart of Picture Processing. It displays many approaches for ordering designs as an example. Photo geometry is concerned with obtaining precise estimates from photographs.

Image Pre-handling: The goal of this interaction is to improve the picture information (highlights) by suppressing undesired contortions and upgrading some key picture elements, so that the PC vision models can benefit from this improved data to work with. Reading the image, Resizing the image, and Data Augmentation (Gray scaling of the image, Reflection, Gaussian Blurring, Histogram, Equalization, Rotation, and Translation) are all steps in the picture pre-processing process[13].

Detection refers to the confinement of an item, which entails the partition of the image and the recognition of the object of interest's location. Highlight extraction and preparation: This is an important step in which factual or deep learning strategies are used to identify the most interesting aspects of the image, including features that may be unique to a particular class and that will later help the model distinguish between different classes. Model preparation is the interaction in which the model takes in the highlights from the dataset. Article grouping: Using an appropriate order technique that contrasts the picture designs and the objective instances, this progression groups distinct articles into preset classes.

III. RESULTS AND DISCUSSION

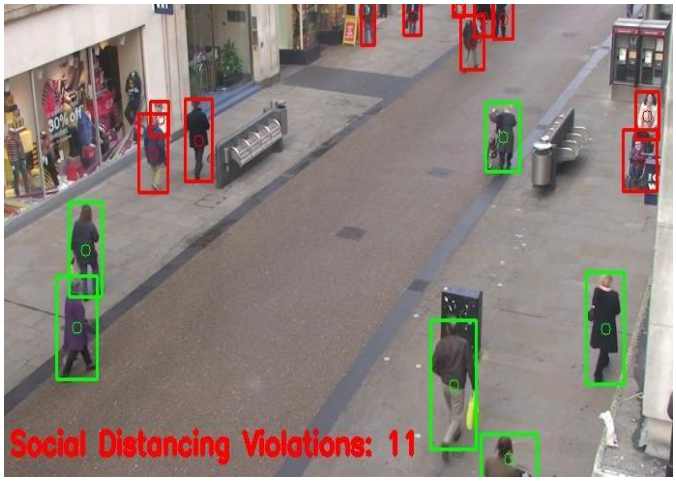


Fig. 4. Video still of result

A pedestrian walks down a public street in the shot. The video frame in this composition is angled towards the street at a specified angle. The video frame's perspectives view is changed to a top-down view for more exact distance estimation. The points represent each pedestrian for the purpose of determining social separation. Red dots represent pedestrians whose distance from another pedestrian is less than the permissible threshold, while green points reflect pedestrians who maintain a safe distance from other pedestrians. Pedestrians approaching another pedestrian too closely until they are overlaid on the camera view may result in detecting errors.

Fig. 5. Increased accuracy of result

The Distance of Each Detected Person from Camera
Depth of Person 0 : 596 cm
Depth of Person 1 : 685 cm
Depth of Person 2 : 704 cm
Depth of Person 3 : 719 cm
Depth of Person 4 : 238 cm
Depth of Person 5 : 1014 cm
Depth of Person 6 : 654 cm

The 3D midpoints (x,y,z) of each detected person in cm
Mid-Points of person : (187, 402, 596)
Mid-Points of person 1: (31, 422, 685)
Mid-Points of person 2 : (1401, 293, 704)
Mid-Points of person 3: (114, 444, 719)
Mid-Points of person 4: (497, 206, 238)
Mid-Points of person 5 : (1869, 187, 1014)
Mid-Points of person 6: (460, 413, 654)

Total PERSONS Detected: 7
Total VIOLATIONS Detected:
28.57142857142857 % of people are NOT maintaining the recommended distance
71.42857142857143 % of people are maintaining the recommended distance
Minimum allowable distance between two persons: 110 cm

The approach used to identify pedestrians has an impact on the accuracy of distance measurements between them. By showing the bounding box, the YOLO method can also detect the pedestrian's half body as an object; however, the pedestrian's position, which corresponds to the middle-point of the bottom line, is approximated using the bounding box, which is less precise. To overcome detection errors, the proposed methodology was changed by adding a quadrilateral box to analyse the defined region in the image. As a result, only pedestrians strolling within the defined space will be registered when people density is measured.

IV. CONCLUSION

The detection of social isolation is done using a deep learning model. Computer vision may be used to determine the distance between people, and any non-compliant pair of people will be highlighted with a red frame and a red line. The proposed method was validated using a video of pedestrians walking down a street. The visualization findings demonstrated that the proposed approach is capable of detecting social distancing measures between people, and that it may be modified for use in a variety of environments, including the office, restaurant, and school. Furthermore, the work may be improved even more by improving the pedestrian detection algorithm, incorporating other detection methods such as mask detection and human body temperature detection, expanding the computing power of the hardware, and calibrating the camera perspective view.

The YOLO algorithm is also able to detect the half body of the pedestrian as an object by showing the bounding box, the position of the pedestrian corresponds the middle-point of bottom line is estimated based on the bounding box will less precise. To overcome the detection errors, the proposed methodology had been improved by adding a quadrilateral box to observe the appointed region in the image. Hence, only the pedestrians walking within the specified space will be counted for people density measurement.

V. REFERENCES

[1]. Suresh K, Adwitiya Mukhopadhyay, "Investigation on COVID-19 by using Machine Learning Techniques", Computing Power and

Communication Technologies (GUCON) 2021 IEEE 4th International Conference on, pp. 1-5, 2021..

[2]. Joel Marvin Tellis, Shivanee Jaiswal, RishikKabra, Parthvi Mehta, "Monitor Social Distance Using Convolutional Neural Network and Image Transformation", Section Signature Conference (IBSSC) 2021 IEEE Bombay, pp. 1-4, 2021.

[3]. Vijaya Shetty, Keerthi Anand, Pooja S, Punnya K A, Priyanka M, "Social Distancing and Face Mask Detection using Deep Learning Models: A Survey", Innovation in Technology (ASIANCON) 2021 Asian Conference on, pp. 1-6, 2021.

[4]. Prerit Gupta, AkshadaGaonkar, Yash Gupta, Dattatray Sawant, "Anti-Covi-Bot: A Robot to Combat COVID-19", Section Signature Conference (IBSSC) 2021 IEEE Bombay, pp. 1-5, 2021.

[5]. I Shiny, Albin Fredy, Jeena Jose, Jeseena.K. P, Rohith P V, "Covid-19 Protocol Enhanced Using Computer Vision System", Innovations in Power and Advanced Computing Technologies (i-PACT) 2021, pp. 1-6, 2021.

[6]. Rakhsith L. A, Anusha K. S, Karthik B. E, Arun Nithish D, Kishore Kumar V, "A Survey on Object Detection Methods in Deep Learning", Electronics and Sustainable Communication Systems (ICESC) 2021 Second International Conference on, pp. 1619-1626, 2021.

[7]. Suresh K, Bhuvan S, Palangappa M B, "Social Distance Identification Using Optimized Faster Region-Based Convolutional Neural Network", Computing Methodologies and Communication (ICCMC) 2021 5th International Conference on, pp. 753-760, 2021.

[8]. Suresh K, Adwitiya Mukhopadhyay, "Investigation on COVID-19 by using Machine Learning Techniques", Computing Power and Communication Technologies (GUCON) 2021 IEEE 4th InternationalConference on, pp. 1-5, 2021.

[9]. YC Hou, MZ Baharuddin, S Yussof and S. Dzulkifly, "Social Distancing Detection with Deep Learning Model".2021

[10]. J. Redmon and A. Farhadi, Yolov3: An Incremental improvement, 2018.

[11]. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[12]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception architecture for computer vision", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826, 2016.

[13]. D.T. Nguyen, W. Li and P.O. Ogunbona, "Human detection from images and videos: A survey", Pattern Recognition, vol. 51, pp. 148-75, 2016.

[14]. A. Krizhevsky, I. Sutskever and G.E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural Information processing systems, pp. 1097-1105, 2012.

[15]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", Computer Vision and Pattern Recognition, 2009.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Reason Behind the Success of Apple

Vijay Selvaraj J, Yashas KM, Teju D, K John Prasad
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

If we were given a choice to buy any smartphone of our choice, irrespective of its cost, most of us would say that we want an iPhone. What has made us think like this? What is the reason behind people preferring Apple products over its other rivals? How did Apple attract such a huge fan base? You will get your answers to all these questions in this paper.

Keywords: smartphone, iphone

I. INTRODUCTION

The company, Apple.Inc[1] was founded in 1976 by two young hackers, Steve Jobs and Steve Wozniak. Its second product, the Apple II, was the first personal computer to achieve mass-market success. The Macintosh, released in 1984, introduced the modern graphical user interface to the mainstream. Apple began to struggle after its board ousted Steve Jobs from the company in 1985. When Jobs returned to Apple in 1997, it was close to bankruptcy. Then Jobs led a spectacular recovery, introducing the iPod in 2001, the iPhone in 2007, and the iPad in 2010. The result: Apple earned almost $40 billion in profits in its 2014 fiscal year. Before submitting your final paper, check that the format conforms to this template. Specifically, check the appearance of the title and author block, the appearance of section headings, document margins, column width, column spacing and other features.

Apple's success is due in large part to its obsessive focus on the user experience. Apple is a designer centric company that likes to build all parts of a product hardware, software, and online services itself. That approach has allowed Apple to build some of the most elegant and user-friendly products ever created. Apple II also became popular in the education market. Apple sold hundreds of thousands of computers to schools that wanted to give their students a chance to learn how to program. Apple continued selling Apple II computers until 1993, when they were discontinued in favor of Macintosh computers. In total, around 5 million Apple II computers were sold.

II. APPLE'S MARKETING STRATEGY

Ever wondered what did Apple do differently to gain such a huge market? One of the main reasons is its marketing strategies[2]. Every new version of a product launched by Apple has a Launch event which is not the case with other brands. All there portray

apple as one of the premium brands and attracts a huge section of the public.

A. Product Placement

Apple has used this strategy to place their products in famous TV Shows and associate it with celebrities, which intrigues the curiosity in consumers to know more about the product. Apple avoids the usual route other companies take to market their products through banner ads and PPC [pay per click] advertisement system.

B. Unique Value Proposition

Rather than competing with similar products on the market in terms of price, Apple uses UVP for their products by offering unique features that simply absent on other devices. Apple has a history of not altering the price of their products until their successors arrive on the market or the product is no more produced.

C. Void Marketing

Applemarkets their products by just ignoring the existence of competing products and just imagines the market is void apart from products from their own line-up. Apple rarely markets their products by comparing their products with other competing products, rather it compares it with its own product from previous generation. Ex., iPhones are always marketed by comparing them to previous iPhones.

D. Simplicity

Everything, right from design to marketing is kept simple and easy to understand. Which helps the user to make conscious decision to choose a product. Apple always names their products for easy understandability of the user. Ex., Naming the iPhones with numbers [and a higher number suggest better product in the same line-up].

E. Speaking the language of consumers

Apple never market their products based on the specifications and hardware used in a product. But they focus on how it creates an impact on the users. Ex., Mentioning the battery capacity is not helpful, while mentioning how long the battery would last is helpful for the user.

F. User Experience

Most of the companies aspire to provide Apple – like user experience to their users. Which itself is sufficient to assess the quality of user experience offered by Apple to its customer.

G. Building a strong locked community

Apple offers exclusive features which only work across their devices within the Apple ecosystem. Users get used to these features and hence hesitate to use or migrate to other platforms. Also, this would force the user of a product from Apple to further invest in their ecosystem. Ex., Apple's iMessage, Airdrop, Handoff work only across iPhones, iPads and Macs.

III. REASONS BEHIND THE SUCCESS OF APPLE

Now, let's bring out the elephant in the box. Why has apple been so successful or rather why is apple successful[3].

A. Employees use Apple products

The people who make an Apple product use it themselves. The creators make products which they want to use and which is easy to use for them. Hence, they make a product which is user friendly for them to use on a day-to-day basis. Whether or not the public preferred it, they used it. Steve represented the real customer. Therefore, all his engineers had to make a product which would please and impress their CEO. They wanted to create a stage where their customer couldn't live without their products.

B. Easy to use

Industrial design is an important aspect which Apple always keeps in mind while making their products. Any user interface which is complex and difficult to understand from the time go is considered worthless

to the consumer. However good the product is, if its difficult to understand and use, the product always has a tendency to fail in the market. Apple realized this from the word go and made sure that whichever product they made, the interface is easy to use, understand and learn. As technology has become much more advanced nowadays and users want more and more features in the products they use, the task of providing these features with an easy interface becomes challenging. However, Apple has somehow found a way to deliver time and again one of the best user interfaces in the market.

C. Keep things simple

If you take the mobile phone industry, apple has only one single product – iPhone. Brands nowadays provide 25-30 models of mobile phones which creates a sense of doubt and confusion in the user's mind. Apple only focuses on only one product at a time and makes sure its users get the best out the only product which they have. Yes, there will be a set of people who do like options and complexity, but research over the past 30 years has proved that choice is nice, but the process of selecting a tech product should be simple. This is how Apple stands out from the rest. Apple makes sure to use the best technology possible in their devices and ensures that their users have the best with them at all times.

D. Excellent customer service

Since the technology that Apple uses is advanced for its users to understand, they provide excellent customer services to help their users with any services the customers want from them. Apple also provides one of the best in-store experiences in the market. No one expected Apple to perform that well in retail but Apple once again made sure it's the best at what it is. This is because the salespersons in the stores know their products really well. They know how exactly treat their customers and get right into the business. Most of the services are taken care of on the spot. This attracts more and more customers. Apple products are easy to understand, but if one has a problem, Apple would never hesitate to take care of it.

E. Apple only makes a product if Apple can do it better

They invent a very few products but make sure they work on the invented products time and again to make it better easy to use it. Apple also makes sure that their technology is at least two years ahead of its time. When apple launches a new iPhone, it will already be working on the iPhone which has to be launched next. The same goes for all its products. As you are reading this paper, Apple is already working on its products which will be launched two years from now

Ever heard people say, "Once you start using an Apple product, you won't feel like using any other products". All the above reasons are why this statement is actually true.

IV. ACKNOWLEDGMENT

I would like to express my gratitude to my reviewer, Dr. Subhashini for her support. I would also like to thank Ms. Surya Pandey for her support and guidance.

V. REFERENCES

[1]. Michael A. Natale and Joshua Kim. "Case Study: Apple.Inc"

[2]. Varun Kapoor, "Apple Swot Analysis, Pest Analysis, Market Analysis, Market Segmentation, Positioning Strategy".

[3]. Abdulla Aljafri, "Apple Inc. Industry Analysis, Business Policy and Strategy".

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Artificial Genetic Evolution with Reinforcement Learning

Harsh Karanwal, Sunandhita B, Vanshika Sai Ramadurgam, Jyothsna N
Department of Computer Science Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper introduces the term Artificial Evolution, here defined as the controlled micromanipulation of genetic information from one generation to the next. Within computer science, artificial evolution is commonly implemented in terms of an evolutionary or genetic algorithm. A GA is a search procedure that implements the three essential elements of Darwinian evolution - replication, mutation, and selection. We also introduce Reinforcement Learning (RL), which is a machine learning method based on rewarding desired behaviors and/or punishing undesired ones. In reinforcement learning, developers devise a method of rewarding desired behaviors and punishing negative behaviors.
We have built a program to demonstrate artificial genetic evolution using reinforcement learning and we have named our project as "Smart Rockets". This is an AI immersed project; it's mainly designed to depict how the phenomena of evolution take place in living and non-living entities. It's represented using 3 major entities, a rocket, an obstacle to obstruct the rocket and target.

Keywords – Genetic algorithm, Reinforcement learning, Artificial genetic evolution, Intelligent retrieval, "Smart Rockets"

I. INTRODUCTION

A. Genetic Evolution and Artificial Genetic Evolution

This Evolution is the change in the characteristics of a species over several generations and it depends on the process of natural selection. Genetic Variation within a population is the backbone of evolution as it affects the characteristics of an organism. This genetic variation comes from mutations in the genome, and reshuffling of genes.
The theory of genetic evolution is based on the idea that all species are related and gradually change over time. The tech industry is expected to grow significantly in the future and this will be made possible using many approaches.
One of the imperative approaches is Artificial Evolution which consists of genetic evolution as the foundation. The relationship between genetic evolution and humans is analogous to that of artificial evolution and robots. Artificial evolution commonly refers to the instantiation of evolution within a non-biological medium.
Artificial evolution is commonly implemented in terms of a genetic algorithm. A genetic algorithm is a search procedure that implements the three essential

elements of Darwinian evolution - replication, mutation, and selection.

B. Reinforcement Learning

Reinforcement Learning (RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences.

RL is similar to supervised learning technique. Both supervised and reinforcement learning use mapping between input and output. In supervised learning, a teacher is assumed to be present during the learning process, when a comparison is made between the networks computed output and the correct expected output, to determine the error, which is then used to change the weight parameters to improve the performance. But in RL, though a teacher is available, it does not present the expected answer but only indicates if the computedoutput is correct or incorrect. The information provided helps the n/w in its learning process. A reward is given for a correct answer computed and a penalty for a wrong answer. Weights are modified in the units that have errors. The goal of RL is to find a suitable action model that would maximize the total cumulative reward. Q-learning and SARSA (State-Action-Reward-State-Action)are two commonly used model-free RL algorithms.

Since, RL requires a lot of data, therefore it is most applicable in domains where simulated data is readily available like gameplay, robotics. RL is quite widely used in building AI for playing computer games. Some other applications of RL include abstractive text summarization engines, learning optimal treatment policies in healthcare and RL based agents for online stock trading. In robotics and industrial automation, RL is used to enable the robot to create an efficient adaptive control system for itself which learns from its own experience and behaviour.

II. METHODOLOGY

A. Objective

- The objective of this project is to demonstrate how evolution takes place in any living or non-living entity.
- It explains how Genetical Evolution is processed by using some major and basic genetic algorithms. Basically it will show and explain how evolution is processed in, say, anything - humans, animals or even machines.
- This project uses machine theory and Java programming language.

B. OOP concepts used

a. Object

An Object can be defined as an instance of a class. An object contains an address and takes up some space in memory. Objects can communicate without knowing the details of each other's data or code. The only necessary thing is the type of message accepted and the type of response returned by the objects.

b. Class

Collection of objects is called class. It is a logical entity. A class can also be defined as a blueprint from which you can create an individual object. Class doesn't consume any space.

c. Data Abstraction

Data abstraction is the process of hiding certain details and showing only essential information to the user. Abstraction can be achieved with either abstract classes or interfaces.

d. Data Encapsulation

Encapsulation in Java is a mechanism of wrapping the data (variables) and code acting on the data (methods) together as a single unit. In encapsulation, the variables of a class will be hidden from other classes, and can be accessed only through the methods of their current class.

III. IMPLEMENTATION

This project has an Artificial Intelligence based interface built with the principles of Neural Networking which is designed in such a way that it depicts and explains how Genetical Evolution is processed by using some major and basic genetic algorithms.

A. Algorithm

1. Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.
2. The rockets are unaware of any obstructions in between and just have to reach the destination.
3. Every phenomenon occurring with a rocket i.e. crashing into the obstacle or the boundaries or touching the destination is sensed and the rocket tries to act accordingly.
4. For example, if a rocket crashes into the obstacle it'll try not to go from that direction again, if it reaches the destination it'll try to follow the same path and improvise it.
5. A minor amount of intelligent retrieval concept can be examined as each rocket acquires the direction to move by the dataset of vectors provided.

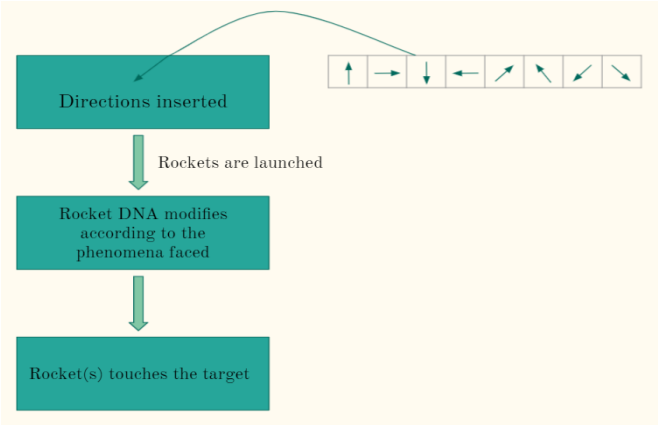


Fig. 1

B. Outputs

In reference to Fig. 2, initially (Generation 1), all rockets are randomly launched in a random direction with each rocket having a unique DNA.

Fig. 2

In reference to Fig. 3, the generation in which the first rocket reaches the target, the DNA of the rockets modifies/changes, hence, their path changes.

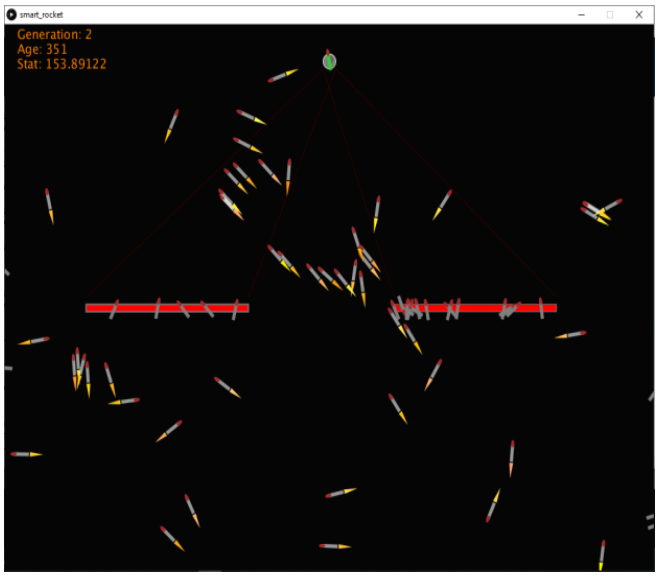


Fig. 3

Then this process repeats several times and each rocket becomes smarter generation by generation. Hence, most of the rockets achieve their path to the

target and the rate of rockets colliding with the obstacles reduces, as shown in Fig. 4.

Fig. 4

Now we can see all rockets have achieved the path to the target with a very low number of rockets crashing into the obstacles.

IV. CONCLUSION

Artificial Genetic Evolution is a theoretical concept which will be implemented in the future along with significant changes in technological trends. In combination with Reinforcement Learning, it will be a compelling topic of interest in the industry.

Smart Rockets is a program built for demonstrating artificial genetic evolution using reinforcement learning. This phenomenon can be represented using the 3 entities concept: a rocket, obstacle, and target, as described in the paper. Over the coming years, the technology industry is expected to grow significantly leading to a decrease in human workforce. Therefore, the main objective of this program is to foster self-learning which is a key factor in accomplishing this goal.

In future, this research will lead to implementing self-learning capabilities in machines and eliminate the requirement of a programmer when the machine needs to be corrected based on its previous experience.

V. REFERENCES

[1]. https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292?gi=479df0807c16

[2]. Artificial evolution of active filters: a case study | IEEE Conference Publication | IEEE Xplore

[3]. https://link.springer.com/referenceworkentry/10.1007/978-1-4419-9863-7_1485

[4]. Paul F Christiano, Jan Leike, Tom Brown, MiljanMartic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In Advances in Neural Information Processing Systems, pp. 4299–4307, 2017.

[5]. Coline Devin, Pieter Abbeel, Trevor Darrell, and Sergey Levine. Deep object-centric representations for generalizable robot learning. In 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 7111–7118. IEEE, 2018.

[6]. Vikash Kumar, Emanuel Todorov, and Sergey Levine. Optimal control with learned local models: Application to dexterous manipulation. In 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 378–383. IEEE, 2016.

[7]. Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. arXiv preprint arXiv:1806.10293, 2018.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Collaborative Learning Using Django Framework and Python Programming

Ajay PV, Ranganath K, Sujey Rajesh Naik
Department of CSE, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

This paper is published by taking into consideration to make learning easy using modern technology. To install application or software, it takes time and consumes storage and decreases the performance of the device. Mainly we don't get all requirements in a single application.Developing different applications takes time and gaining more user for developed apps is more difficult. mobile and computer applications are bulkier than web apps.

Power store is a collection of programs where each program is considered as app so basically it is a collection of apps where users can share their apps to public and other people get to know multiple shortcuts to implement the code using the source code of the apps shared by the other developers .It is totally web based there and is no need to install apps and code will be loaded on time so it will be super-fast. User must create a power store account or authenticate to login then user can create an app or view other users apps and even use others apps functionality which is shared by the other coders and developers.

Keywords: Research Paper, Technical Writing, Django, WEBRTC and Google cloud

I. INTRODUCTION

Basically, this Research Paper contains the idea that was been implemented as a project by our team using Django framework, which is built top on python, google cloud, webrtc, firebase, python, Html, CSS, JavaScript and other python libraries. This project is collection of small projects, which is considered as a web app, each web app has different functionality. where every user can access just using the link or by searching on the internet. In this project every new developer get exposure to the new way of writing code which are already coded by the experience developer.

If any user or developer gets any new idea, he/she can convert his idea and code and upload in this project. This code will be converted to a web app.

They can write in languages like html, CSS and JavaScript. We built some of the apps in this project which helps the students to increase their communication skills and technical skills. There are so many students having good knowledge but lack of communication skills and even some people find any new concept to learn, it may take most of their time to learn by them. Self-learned concept will not be remembered for a long time and even bad communication skills leads to the less connection in

our network and some people can't express what exactly they want to express.

To solve this kind of problems, we build few apps in the project considering the above reason which will be helpful for building their communication skills by communicating with other user by face-to-face communication. Thanks to WEBRTC technology which is helpful in transferring real-time data between the communicating users and even users can share their screen to get the help from the other users for the problems that cannot be solved by them where helpers get their rewards.

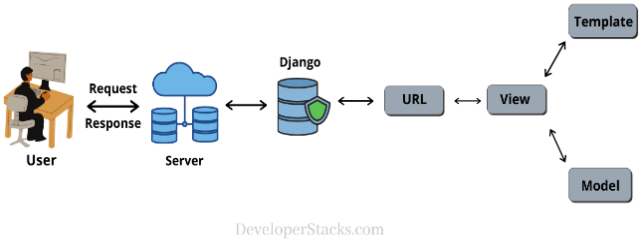
In this project the data is not handled by the server, instead the data communication takes place between the two communicating peers. In this part signalling is handled by the Firebase which is part of google cloud using some free stun servers provided by the google.

For securing user data google API is used for authentication purpose. There is one more app that helps in creating notes where every user can contribute their ideas to the notes.

Apart from the apps mentioned above we have also included an app that entertains people using videos and images for some fun in the life. We also used the git version control system to build this project at the faster rate. So, this project is a combination of all the things that is learning, building communication, building technical skills, enjoying with all these skills. "This is learned with fun app".

II. METHODS AND MATERIAL

Django framework – Django is a python web framework that helps in speedy blooming of secure and maintainable websites. [3]Django handles much of the fuss of web development, so that user can focus on building their app without needing to recast the wheel. It is free and open source and has a flourishing and mobile community, great documentation, and many more options for support. Everything that is required is made a part of one product and it all functions seamlessly together, also follow consistent design principles, and has substantial and latest documentation



Django assists developers avoid many common security mistakes by providing a framework that has been engineered to do the right things to protect the website. Django helps connects python code and html dynamic content and data stored in MySQL.

Python language-Basic language using which the project is mainly built. Python is an interpreted, object-oriented, high-level programming language which was developed by Guido van Rossum. It was designed to be easy as well as fun. Python is a beginner-friendly language, replacing Java and other programming languages as the most widely used language because it handles much of the complexity. It allows beginners to focus on understanding programming concepts rather than minute details. Python is usually used for server-side web development, software development, mathematics, and system scripting, and is popular for Application Development and as a scripting language with built-in data structures, dynamic typing, and dynamic binding. Maintenance costs for the program is reduced with python because of easy to learn syntax and prominence on readability. Python support for modules and packages allows modifying programs and reuse of code. Python is an open-source language, so

innumerable independent programmers are building libraries and functionality for it.

Bootstrap- To add the colorful content to HTML code by animating content. Bootstrap is sleek, intuitive front-end framework for faster and easier web development. It includes HTML and CSS based design templates for creating common user interface components like forms, buttons, navigations, dropdowns, alerts, modals, tabs, accordions, carousels, tooltips, and so on. Bootstrap gives user the ability to create flexible and responsive web layouts with less efforts. [4]Bootstrap 3, framework consists of Mobile first styles throughout the entire library instead them of in separate files. It is supported by all popular browsers.

With just the knowledge of HTML and CSS anyone can get started with Bootstrap. Also, the Bootstrap official site has a good documentation. Bootstrap's responsive CSS adjusts to Desktops, Tablets and Mobiles. More about the responsive design is in the chapter Bootstrap Responsive Design. It offers a clean and uniform solution for building an interface for developers. It contains beautiful and functional built-in components which are easy to customize. It also provides web-based customization, and it is an open source.

HTML, CSS, JavaScript- Used to design frontend part. Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript are the languages that run the web. They're very closely related, but they're also designed for very specific tasks. Understanding how they interact will go a long way towards becoming a web developer. HTML, CSS, and JavaScript are the coding languages. These languages are used to build websites and web applications. HTML is the base of every website. It is impossible to build a website without using the HTML codes.

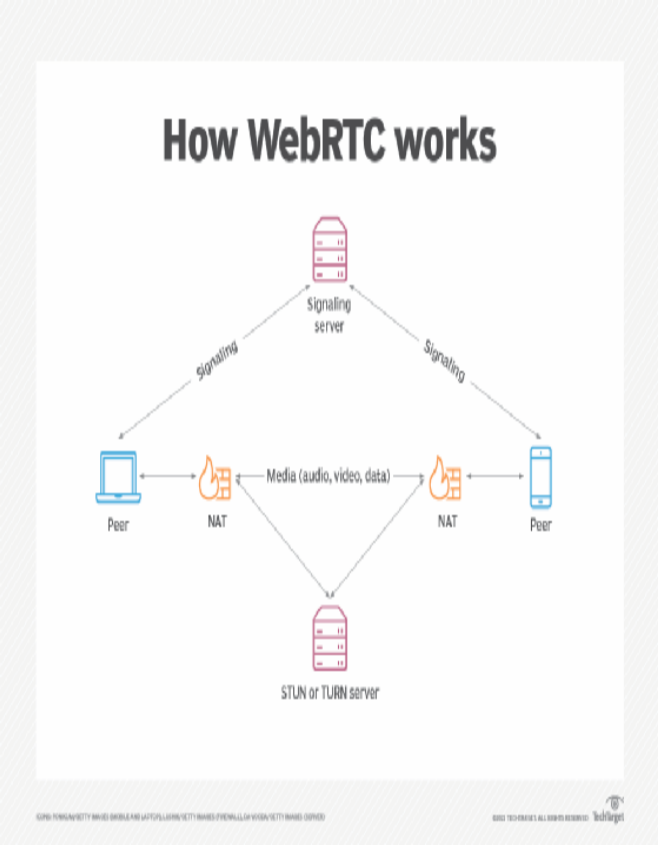
HTML stands for Hyper-Text Markup Language. These are used to build the structure of the website. It has the skeleton of the website body. CSS stands for Cascading Style Sheets. It works as the makeup for the HTML. CSS improves layout of a website structure of the website and do other jobs to build the impressive graphics. JavaScript is a programming language that is used to add interactivity and functionality to a website.

Amazon web services- For storing static files in the cloud. Amazon web service is a platform that provides scalable and cost-effective cloud computing solutions. AWS is a widely adopted cloud platform that helps several on-demand operations like compute power, database storage, content delivery, etc., to help corporates scale and grow's helps businesses to build several complex applications. Organizations of every size can run all imaginable use case on AWS.

[2] AWS is used by many businesses because it offers multiple types of storage to choose from and is easily accessible as well. AWS is used for storage and indexing and to run critical business applications. It has made possible for the businesses to host their websites on the AWS cloud, like other web applications. A lot of computing power needed to run gaming applications. AWS does it easier to enable the best online gaming experience to gamers. A feature that separates AWS from other cloud services is its

ability to launch and scale mobile, e-commerce, and SaaS applications. API-driven code on AWS can enable companies to build complex scalable applications without requiring any OS and other systems.

Firebase and WEBRTC API's – Thisis used for signaling. To share the real time data. [1]webRTC stands for web real time communication. The way this works is called peer-to-peer connection between the browser with a signaling state which is done by the server. WebRTC is a definition that enables browsers, phone, and native clients to exchange video, audio, and general information with the help of API's.



With the help of this technology, communication is usually between peer-to-peer and direct. This allows for easy approach to media device on hardware technology. This works by establishing a connection between different peers with the help of signaling, which then facilitates streaming or exchanging of data. It is important to know that these technologies are executed as an open web standard or specification and are exposed with JavaScript APIs in all major browsers.

Google API's- To authenticate user with the google account. Google APIs makes use of OAuth 2.0 protocol for authentication and authorization. [5]Google supports common OAuth 2.0 scenarios such as those for web server, client-side, installed, and limited-input device applications. Obtain OAuth 2.0 client credentials from the Google API Console. Then the client application requests an access token from the Google Authorization Server and then extracts a token from the response and then sends the token to the Google API that you want to access. For an interactive demonstration of using OAuth 2.0 with Google including the option to use your own client credentials, experiment with the OAuth 2.0 Playground.

MySQL- To store the user data. MySQL is an open-source, fast reliable, and flexible relational database management system that is typically used with PHP. MySQL is a database system used for developing web-based software applications. MySQL used for both small and large applications. MySQL is easy to use. MySQL supports standard SQL (Structured Query Language). MySQL is free to download and use. MySQL is presently developed, distributed, and supported by Oracle Corporation.

MySQL is Written in C, C++. A database is a separate application that stores large amount of data. Each database contains distinct APIs for creating, accessing, managing, searching and replicating the data present in it. Other kinds of data stores such as files on the file system or large hash tables in memory can also be used but data fetching and writing would not be so fast and easy with such systems. Relational database management systems (RDBMS) are used to store and manage huge volume of data. All the data is stored

into different tables and relations are established using primary keys and Foreign Keys.

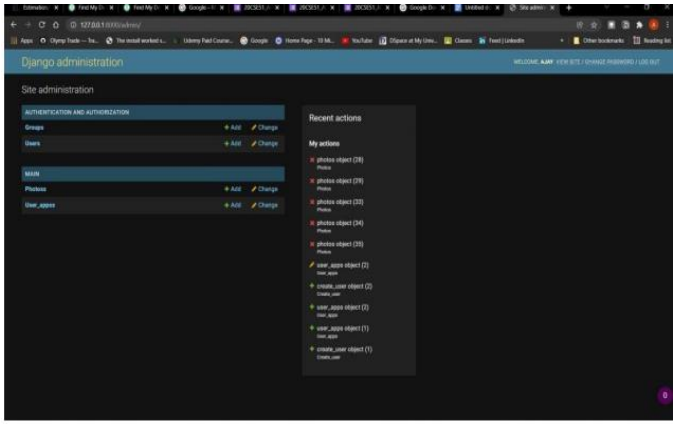
Git- For version control of the project. Git is a version control system that developers use all over the world. It helps users track different versions of their code and collaborate with other developers. If the user is working on a project over time, then they may want to keep track of which changes were made, and by whom, and when those changes were made. This becomes increasingly important if users end up having a bug in their code! Git can help with this. GIT is a version control system that needs to be downloaded onto your computer. It is essential to use Git if user wants to collaborate with other developers on a coding project or work on your own project.

III. RESULTS AND DISCUSSION

All paragraphs must be indented. All paragraphs must be justified, i.e., both left-justified and right-justified.

Admin Functionality

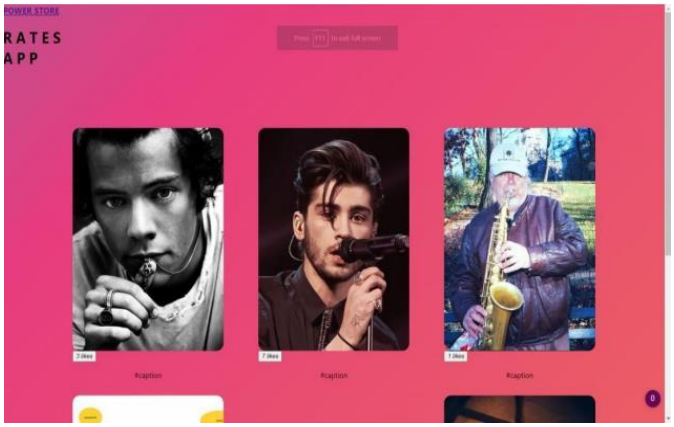
Admin is the one who controls the entire application. Functionalities available to admin are:



In admin option, admin can add new social authentication option to the project. admin can get number of users using project website, admin can't see the users' passwords it is totally encoded.

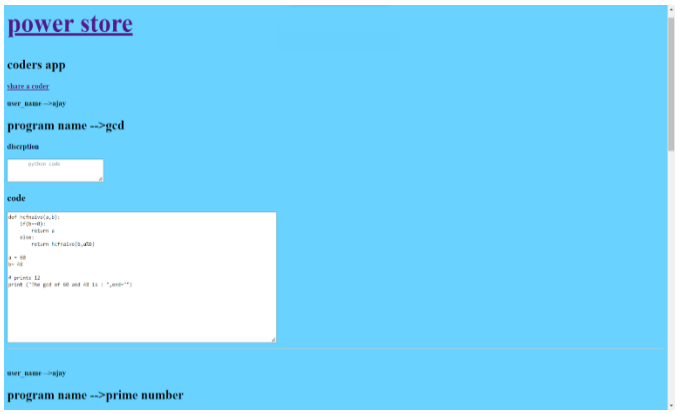
Power Rates

This app prebuilds in power store to get idea to other uses and to gain more users for the project. it is basically getting the photos and gifts from the users and share to other uses to show the users art talent



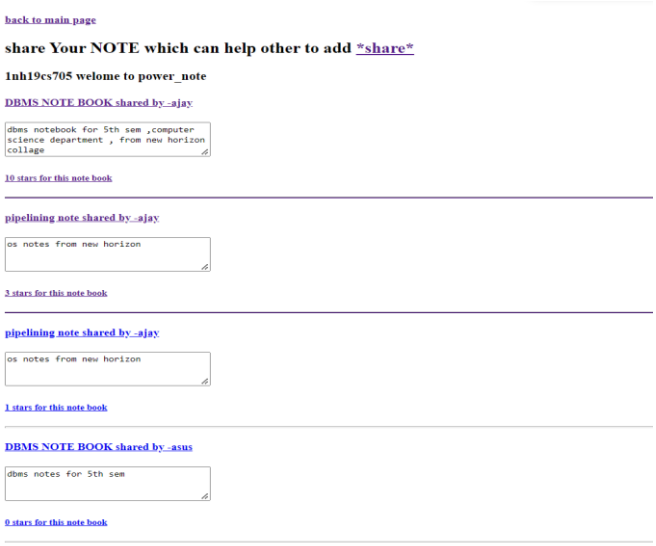
Power Coders App

In this app user can share new ideas which is implemented in code, like implementing a code from math problems which learnt in math class, but to implementation must in html, CSS and JavaScript.



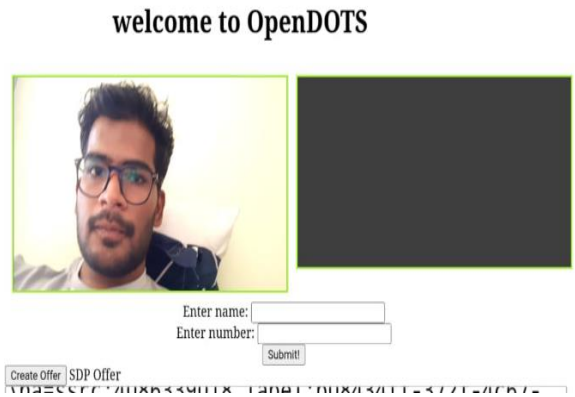
Power Note

In this app user can share notes to another user. if user is created good notes in any subject so they can share their notes to other users so other users learn in essay way. if the user shared notes more like it will in the top of the list.



OpenDOTS

In this app users can lean new concepts by discussing to each other using webrtc face to face video communication. And new user is connected.To the other uses and helping to make new friends and increase their communication skills.



IV. CONCLUSION

The Power store is a collection of programs where each program is considered as app so basically it is a collection of apps where users can share their apps to public another people get to find multiple short ways to implement the code. it is totally web based there is no need install apps and code will load on time so it will be super-fast. 'User must create a power store account or authenticate to login then user can create an app or view other users' apps

V. REFERENCES

[1]. Zafran M RM, Gunathunga L G K M, Rangadhari M I T, Gunathne M D D J, Kuragala K R S C B, MrDhishanDhammearatchi, "Real Time Information and communication Center based on WebRTC", international journal of science and research Publication, volume 6, Issue 4, ISSN 2250- 3153,April 2014.

[2]. Mosca, P., Zhang, Y., Xiao, Z. and Wang, Y., 2014. Cloud security: Services, risks, and a case study on amazon cloud services. Int'l J. of Communications, Network and System Sciences

[3]. Holovaty, A., & Kaplan-Moss, J. (2008). The Definitive Guide to Django: Web Development done right. Apress.

[4]. Faraway, J. and Jhun, M. (1990) Bootstrap choice of bandwidth for density estimation. J. Amer. Statist. Assoc

[5]. S. Challita, F. Paraiso, and P. Merle, "Towards Formal-based Semantic Interoperability in Multi-Clouds: The fclouds Framework," in the 10th International Conference on Cloud Computing (CLOUD). IEEE, 2017,

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Comparative Survey on The Classification of Machine Learning Algorithms

Rakshitha B A, Sharanya M, Soundarya S M
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Machine learning offers the potential for effective and efficient classification of remotely sensed imagery. The strengths of machine learning include the capacity to handle data of high dimensionality and to map the classes with very complex characteristics. Nevertheless, implementing a machine-learning classification is not straightforward. This article therefore provides an overview of machine learning from a theoretical perspective. This article provides classification ofmachine learning algorithms as supervised, unsupervised and reinforcement. Issues considered include the choice of algorithm, selection and optimization.

Keywords: Machine Learning, Supervised, Unsupervised, Reinforcement.

I. INTRODUCTION

Machine Learning is one of the fastest growing fields in today's world. Earlier companies used large data centres to store data and hence the retrieval and management of data was hard. It required enormous human and machine effort. It was tough to generate new data based on the previous data as well. But now-a- days companies make use of the Machine Learning technology to maintain and evaluate their data.[1] With Machine Learning it becomes a lot easier to store, retrieve and manage data. The process becomes faster. Machine Learning is found at every nook and corner of the world. From the face lock feature on our mobile phone to the recommendation systems on various applications like Netflix, Amazon, etc. Machine Learning is used everywhere. Machine Learning is not just a Database problem but also an Artificial Intelligence (AI) problem. Machine Learning is used not just to store previous data but also to generate new data. The idea of Machine Learning is pretty straight forward. We explore and prepare the data for use, train our machine to work with data, evaluate the model, improve the model and deploy it. We make a list of features that we require and set a target to be obtained. Machine Learning models may be descriptive (gaining knowledge from data) or predictive (predicting the future data).

Machine Learning is a vast ocean. It has many ideologies and methodologies. Everyday there are so many new innovations made in this field. Here in this paper we discuss the classification of algorithms in Machine Learning. The algorithms in Machine Learning are classified as supervised learning, unsupervised learning and reinforcement learning.[2]

II. SUPERVISED LEARNING

Let us consider a problem of predicting whether a patient has covid or not. The basic idea behind the solution would be to classify people based on their symptoms. If a patient has symptoms like cough, fever or breathing difficulty then the he/she might have covid else he/she may not (neglecting the asymptotic patients if any). This solution approach can be implemented using Machine Learning. We will have two sets of data to be fed into the machine. One set would have the symptoms of previously infected patients and another set would have health conditions of patients who tested negative. The technical term for this data is datasets. These labelled datasets are to be fed into the machine and a model is to be prepared. We can further use this model to test the possibility of covid infection in the future set of patients. This approach of training a machine based on labelled datasets is called supervised learning. It can be defined as learning a class from its positive and negative examples. Supervised learning can be further classified as Classification and Regression.

Classification

Consider the previously discussed covid problem. Every patient will be categorised into one out of the two categories. A patient may either be infected by covid or not. This is classification. As the name suggests each input will be classified into one of the possible categories. The number of categories starts from a minimum of two and can go up to any number. Algorithms under classification can be further categorised as binary (two classes), multiclass (multiple classes) and multilabel (where an input could belong to more than one class) classification. In this section we will be discussing a couple of algorithms under classification.

Stochastic Gradient Descent Algorithm

The Stochastic Gradient Descent (SGD) algorithm is a simple but efficient Machine Learning algorithm. It is an optimizing algorithm whose approach is to fit the linear classifiers and regressors under convex loss functions. [4] SGD has been around since almost 1960 but has been widely used for large scale learning since only recently. It is extensively used in applications like classification of text, processing of natural language and in representation of graphical models. [6] It finds the model parameters that correspond best fit between predicted and actual output. SGD may be inexact but it is very powerful. [5] Setting the right learning rate of an SGD based solution is of prime importance.

1. Orthogonal Matching Pursuit

. The Orthogonal Matching Pursuit algorithm is used for analysing the fit of linear model with constraints imposed on the non-zero coefficients. It is based on the greedy algorithm which includes determining at each step the atom highly correlated to the current residual. It is quite the same as the Matching Pursuit (MP) method, but it is much better than MP. Duringeach iteration, the residual is recalculated using an orthogonal projection on the space of the previously chosen dictionary elements. [4]

A. Regression

- Regression is a category of supervised learning where the results cannot be categorized into classes unlike classification. Regression is a method for determine the relationship between independent variables (also called features) and dependent variables (also called outcomes). As mentioned earlier no classes or categories can be created. The values predicted are usually continuous. Consider the problem of predicting the rent of a house. We can predict the rent of a house but cannot categorize the rent into classes. In regression we feed the dataset into the machine and it tries to map the features to the outcome. Hence the relationship between the two is determined. In this section we discuss a couple of algorithms under regression. [7]

. **Quantile Regression**

Quantile Regression is a methodology under regression where we calculate the median or other quantiles of y based on x. Quantile means collection of points in distribution that relate to an arrangement of values based on rank in that distribution. Quantile Regression allows us to understand the relationship between variables outside of the mean of the data. This helps in understanding the outcomes non-normally distributed and that has non-linear relationships with the features. Most implementations of quantile regression are based on the linear programming problem. [8]

Least Angle Regression

Least Angle Regression is an algorithm used for high dimension data. It is very similar to the forward step wise regression. On every stage it detects the feature most correlated to the target. If two such features are found that are equally correlated to the target then, instead of proceeding with the same feature it moves forward in a direction equiangular to both the features. The algorithm can be easily modified to produce efficient solutions to other estimators too. [4]

III. UNSUPERVISED LEARNING

Unsupervised learning requires training a model without labels being available toguide the model. Consider the example, imagine that you have a large set of compound-target connected and required to sort them into active and inactive categories. In a supervised approach,every compound-target association will be labeled as either active or inactive. However, without a label, they won't be able to tell if a molecule is active or inactive based on its qualities. Clustering is a term for this. Clustering is a critical component of unsupervised learning.

[11]The compounds' associations with distinct targets, or the qualities that indicate a compound's activity on a target of a given target class, may make this more difficult.Compounds can be correlated with multiple targets and target classes at the same time, making this clusteringdifficult to completely comprehend. Furthermore, it is frequently uncertain how to effectively explicate the proposed grouping; clustering interpretation is frequently ambiguous.

Muchstudypresupposes that clustersare seen asunbiased and significant, while ignoring howor what constitutes “meaningful” in the first place. Furthermore, the chosen approach and its interaction with the properties of a given dataset have a significant impact on cluster outcomes, making it possible to detect clusters in homogeneous or random datasets data. In conclusion, every dataset has clusters, which are revealed using various approaches, but not all of these clusters are interesting or even real.[11]Unsupervised learning can be classified as Clustering and Association.

Clustering

Clustering is an unsupervised machine learning job that splits data into clusters, or groups of related items, without the need for human intervention. It accomplishes this without being told in advance how the groups should appear.[12] Clustering is utilised for knowledge discovery rather than prediction because we may not even know what we're looking for. It elucidates the natural groupings that can be found in data.

Affinity Propagation

Affinity Propagation generate clusters by directing messages between data points until convergence. Unlike clustering algorithms defined as k-means or k-medoids, affinity propagation does not depend on the number of clusters to be known or estimated prior to running the algorithm, for this reason the two important parameters that are to be preferred arewhich controls how many exemplars (or prototypes) are used, andthe dampingfactor which dampens the influence and availability of messages, therefore avoiding numerical oscillations while updating these messages.

[13]A dataset is outlined by a trivial number of eamples, which are input set members who are representative of clusters.The communications

between pairs define one sample's suitability to serve as an example for another, which can be updated in response to the values from other pairings. This update occurs repeatedly until convergence, at this point the final examples are selected, and therefore the last clustering is obtained.[4]

Mean Shift

Mean Shift is an unsupervised clustering approach that seeks to find blobs in a dense sample set. It's a centroid-based method that works by updating centroids candidates to be the mean of the points in a given area. In order to construct the final collection of centroids, these candidates are filtered in a post-processing stage to remove near-duplicates.[14] The difference between K-Means algorithm and Mean-Shift is that latterwon'trequirethe number of clusters to be specified in advance because the number of clusters are to be decided based on data.[15]

Association

The unsupervised learning method of association rule learning (ARL) is used to uncover relationships between variables in huge databases. ARL can handle non-numeric data points, unlike certain machine learning methods. In a nutshell, ARL is concerned with determining how particular variables are related to one another. [10]Association rule is descriptive not the predictive method, generally used to discover interesting relationship hidden in large datasets.People who purchase a motorcycle, for example, are more likely to get a helmet. It can be profitable to find such connections. If customers who buy Product X also buy Product Y, for example, an online shop can suggest Product Y to anyone who is buying Product X.

[10] If/then sentences are at the heart of association rule learning. These remarks can highlight connections between disparate pieces of information. Support and confidence are also used to observe the if/then patterns or linkages. The frequency with which the if/then connection appears in the database is specified by support.Support specifies how often the if/then relationship appears in the database. Confidence defines the number of times the if/then relationship was found to be valid.Marketbasketanalysis and webusagemining are made possible with the association rule.[10]

Apriori Algorithm

Apriori algorithm is a associative learning algorithm which is generally used in data mining. It follows the principle that people who bought this will also buy this. It analyzes the data present in database and extend the number of data items present in that record. It determines the certain amount of association rules which used to determine the trend in data and based on that data items are added or extended.[9]

Its most used case is Market Basket Analysis. It is beneficial for both customer and seller. It helps customer to buy thing and also increase the sales for seller. It also been used in health care for detection of adverse drug reaction. It also used in recommendation system as it learns from the trend according to user preference and recommend data based on that.[9]

ECLAT algorithm

Equivalence Class Clustering and bottom-up Lattice Traversal(ECLAT) is an data mining algorithm used to achieve itemset mining and locate frequent items. Apriori algorithm utilizes a horizontal data format, it must scan the database numerous times to identify frequent items. On the other hand, ECLAT takes a vertical approach and is generally faster as it needs to scan the database only once.[10]

Frequent pattern (FP) growth algorithm

The frequent pattern (FP) growth algorithm is an enhance version of Apriori algorithm. This algorithm represents the database in the form of a tree structure known as a frequent tree or pattern.Such a frequent tree is most likely used for mining the most frequent patterns. While the Apriori algorithm needs to scan the database n+1 times (where n is the length of the longest model), whereas the FP-growth algorithm requires just two scans of the database.[10]

IV. REINFORCEMENT LEARNING

Reinforcement Learning has always been a part of Artificial Intelligence and it is a perfect tool in building artificially intelligent systems and solving the sequential decision making problems. Reinforcement Learning has taken many credits for the breakthroughs in the recent years and it is able to reach a higher level than the human level in many fields. [16]

Reinforcement Learning is a way of learning throughthe interaction with in an environment by taking various actions and facing many successes and failures while trying to maximize the received target required(rewards). The machine(agent) is not informed about which action to be taken. Reinforcement Learningcan be compared to the natural learning processes where in a supervisoror a teacher will not be present and learning process evolves with trial and error techniques, different from supervised learning, in which a machine(agent) needs to be told what the correct action is to be taken for every situation it encounters. [16]

- Reinforcement Learning overlaps with multiple fields: Computer Science, Neuroscience, Engineering, Mathematics, Psychology, and Economics. Figure 1 demonstrates these intersections.

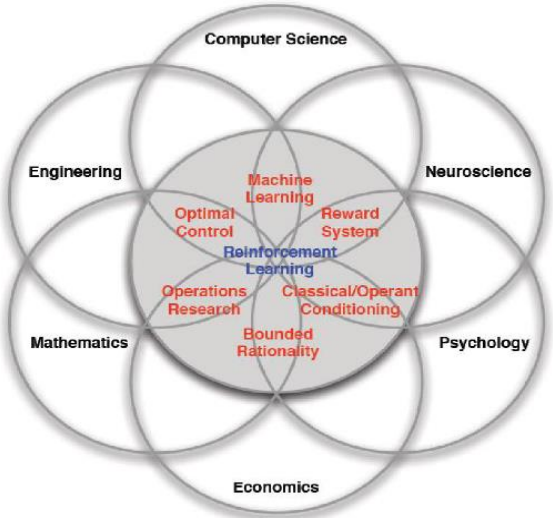


Fig. 1 Reinforcement Learning faces

. Standard Model of Reinforcement Learning

- ` The main terminologiesused in Reinforcement Learning system are: Agent(A), Environment(e), Reward(r), State(s), Policy(***π***), Value(V), Value Function.[18]

- The Agent(A) is a body or an entity which demonstrate various actions in an environment to get rewards.
- The Environment(e) is a backgroundwhere an agent needs to perform actions.
- The State(s) is a latest scenario returned by an environment.
- The Reward(R) is an instant gain that is earned by an agent when it demonstrates some actions. The reward is either positive, negative or null.
- The Policy(***π***) is a blueprint of actions that is applied by an agent to go with the next action depending on the present state.
- The Value(V) is an anticipated long-term yield with a discount, when compared to the small reward.
- The Value Function gives the value of the state and is the calculated amount of the reward.

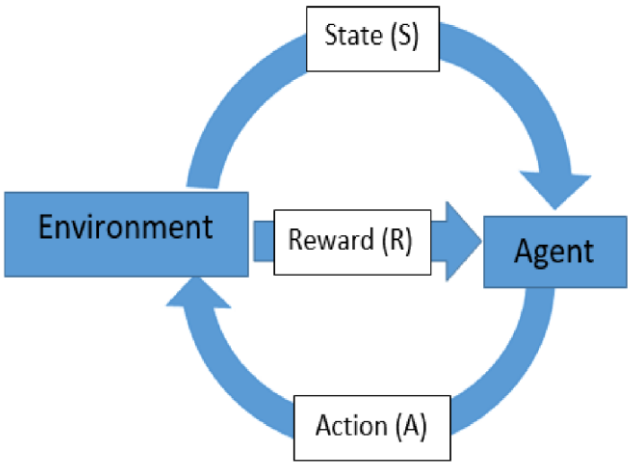


Fig. 2 Reinforcement Learning standard diagram

B. A Simple Example-Tic Tac-Toe

To emphasize on the basic concept of theReinforcement Learning, examine the below example [17]

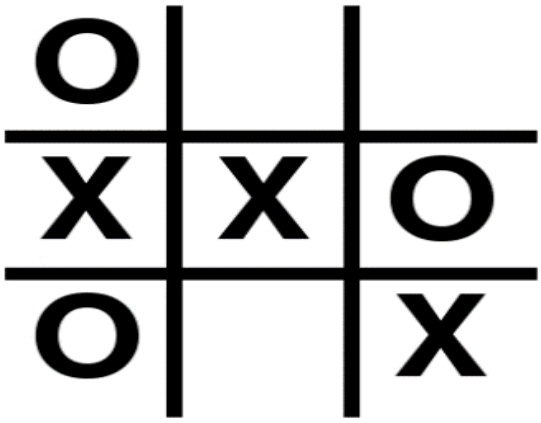


Fig. 3 Tic-Tac-Toe

Here the tic-tac-toe problem is approached with a method which makes use of a value function. Initially, it is needed with setting up of a table of numbers, each one for a possible state in the game. Every number is going be the newest sum of the likelihood of winning from that state. This total is treated as the state's value, and the entire table will be the learned value function. The comparison between the two states A and B will be that, if thestate A has a higher value than the state B,then the present sum of the likelihood of winning from state A is greater than that ofstate B. Supposing thattheirwillalways be a play of X's, then the states with the three X's in a row will have a likelihood of winning of 1, because 1 is considered as a win. Similarly, the states with the three O's in a row, has a likelihood of getting a 0, as it is considered as a failure. Setting the starting values of the remaining states as 0.5, which represents aassumption that their will be a 50% chance of winning.

To select the moves, the agent needs to inspect the states that will be an outcome from each of the practicable moves (one from each empty space on the board) and pick up the present values in the table. Often it is necessary to move selfishly, by selecting a move that will lead to the state with the highest value, which is the state with the highest sum oflikelihood of winning from that state. Rarely, moves can be selected from the left over moves instead. These moves are called as the exploratory moves because they go through the states that can never be seen otherwise.Figure 4 illustrates a series of moves made and which are considered during a game. [17]

There is a try givento make the states give more accurate sum of the likelihood of winning. To do this, theapply the "back up" which is to move the state before each greedy move after a greedy move is done, as shown by the arrows in Figure 4. More accurately, the present value of the previous state's value is madeas close as to the value of the future state's. This can beaccomplished by moving the previous state's value to a tiny part of theso as to near the value of the future state's. Considering the state(s) prior to the greedy move, and s', the state following the greedy move, the modification towards the calculated value of the s, denoted by V(s), can be interpreted as $V(s) \leftarrow V(s) + \alpha [V(s') - V(s)]$ where α is a tiny positive amount called the step-size parameter, that impacts the learning rate. This modificationstep is an example of a temporal-difference learning method, it is called as that because the changes in it are based on a difference, V(s')–V(s), between the calculated estimates at two different times. [17]

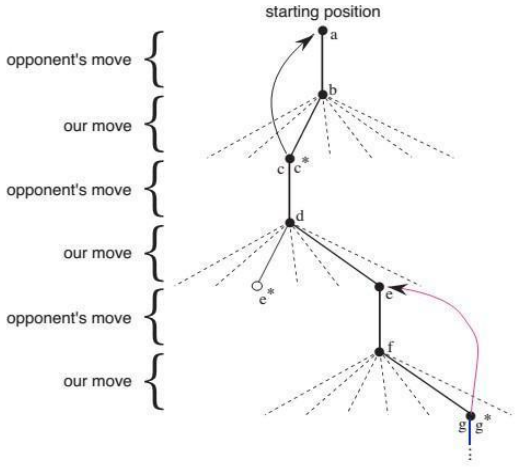


Fig. 4 A series of tic-tac-toe moves

C.Markov Decision Process(MDP), The Basic Model

The mathematical approach of illustrating a resultin the field of Reinforcement Learning isthe Markov Decision Process or (MDP). The Reinforcement Learning problem is a simple orientation of the problem of gaining knowledge from arelation to achieve a goal. The entity that

interacts with the environment and the decision maker is called an agent. The thing the agent interacts with, consisting of the area outside it, is called as the environment. The agent and the environment interact at each of a series of distinct time steps, t = 0, 1, 2, 3, ... At each and every time step t, the agent gains some characterization of the environment's state, $S_t \in S$, where S is the group of possible states, and on that criteria,the agent chooses an action, $A_t \in A(S_t)$, where $A(S_t)$ is a group of actions available in the state S_t. [20]For the next step, in consideration as a result of its actions, the agent gains a numerical reward, $R_{t+1} \in R \subset R$, and emergesbeing in a new state, S_{t+1}.

The agent implicates a map through from the states to the likelihood of choosing from each possible action.

. Markov Property

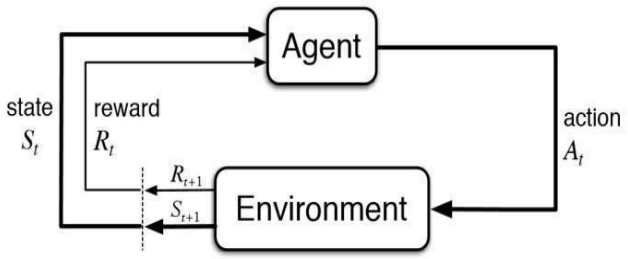
A property that needs to be satisfied by a Reinforcement Learning task to be called as a Markov Decision Process is called the Markov property. Assume that there are finite number of states and reward values. Regarding how a common environment might respond at time t + 1 for the action done at timet. In the most basic case, this response may pivots on everything that hadoccurred earlier. In this case, the results can be defined only by identifying the complete joint probability distribution:

$Pr\{S_{t+1} = s', R_{t+1} = r \mid S_0, A_0, R_1, \ldots\ldots, S_{t-1}, A_{t-1}, R_t, S_t, A_t\}$, [17] for all r, s', and all the feasible values of the past events: $S_0, A_0, R_1, \ldots, S_{t-1}, A_{t-1}, R_t, S_t, A_t$. If a state's signal follows the Markov property, then the environment's feedback at t+1 depends only on the state and the action illustrations at t, in which case the environment's result can be defined by specifying only

$p(s', r|s, a) = Pr\{S_{t+1} = s' , R_{t+1} = r \mid S_t = s, A_t = a\}$, [17]

for all r, s', s, and a. In other words, a state signal follows the Markov property, and will be called as a Markov state, if and only if is equal to $p(s', r|S_t, A_t)$ for all the states s', r and histories, $S_0, A_0, R_1, \ldots, S_{t-1}, A_{t-1}, R_t, S_t, A_t$. In this scenario, the environment and the task as a bundle will have the Markov property. [17]

his mapping is called as the agent's policy and is indicated as π_t, where $\pi_t(a|s)$ is the likelihood that $A_t = a$ if $S_t = s$. Reinforcement Learning techniques specifies how an agent modifies its policy with the outcomes of its experience. The agent's target,is to enhance the total amount of reward it receives to a maximum over a long period of run. [17], [21]



Finite Markov Decision Process(finite MDP)

A task in Reinforcement Learning which satisfies the Markov property is called the Markov decision process, or MDP. If the action spaces and states are limited, then it is called as the finite Markov decision process (finite MDP). [17]

A finite MDP is illustrated by the properties named the state and the action sets and by the one-step results of the environment. Stated any action and the state a and s, the likelihood of each feasible pair of upcoming state and reward, s', r, is denoted $p(s', r|s, a) = Pr\{S_{t+1} = s' , R_{t+1} = r \mid S_t = s, A_t = a\}$.

One can find anything they would want to know about the environment, such as the anticipated rewards for the state–action pairs,the state-transition probabilities and the anticipated rewards for the state–action–next-state triples.

Reinforcement Learning Algorithms for Control Learning

In Reinforcement Learning, three approaches are employed to implement an algorithm. [18]

i. Value-Based

In the abovementioned reinforcement learning method, it's target is to enhance and make the value function V(s) maximum. In this procedure, the agent is expects a long-term reward of the

present states which is under the implication of the policy π. [18]

i. Policy-Based

In this Reinforcement Learning method, the target is to get a policy that when an action is demonstrated in each and every state will help to obtain thehighest reward in the future.

Two variations in theabove mentioned method are:

- In thisReinforcement Learning method, the necessity is to create anenvironmenttypeof virtual model. The agent becomes competent to demonstrate in that mentioned environment.

. Reinforcement Learning Algorithms

There are various algorithms present which solves the Reinforcement Learning problems and two of them are mentioned here

Monte Carlo Learning

Monte Carlo learning method is a very simple concept where the agent gains an understanding about the states and rewards when it interconnects with the environment. In this method, the agent createsthe experienced samples and then based on the aggregate returns, the value is estimated for a state or a state-action. Below are the key characteristics of Monte Carlo(MC) method: [22]

- There is no model.
- Agent learns from a sampled experience.
- Learns the state value vπ(s) under policy π by experiencing the average return from all the sampled episodes (value = average return).
- Only after a complete episode, values are updated (because of this algorithm convergence is slow and update happens only after an episode is completed).
- There is no bootstrapping.
- Only can be used in episodic problems.[22]

Q-learning

Q-learning is an Reinforcement Learning algorithm which is model-free where the agent learns the value of an action in a specific state. It can handle the problems with stochastic transitions and rewards without requiring any adaptations. [19], [23]

Deterministic: For any state, the produced action by the policy $\boldsymbol{\pi}$ is the same.

- Stochastic: Every action posses a particular likelihood, which is found by the following equation.

Stochastic Policy:n{a\s) = P\A, = a\S, =S] [18]

V. CONCLUSION

We have seen so far the different learning methodology under Machine Learning. We have seen 3 types of learning methodologies under Machine Learning supervised learning that uses labelled datasets, unsupervised learning that uses unlabelled datasets and reinforcement learning that looks for hidden patterns in data.

We have also seen algorithms under each type. As already mentioned, Machine Learning is a vast ocean where new things are being discovered every now and then. The points discussed in this paper comprise of a small subset of Machine Learning.

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	The machine learns by using the labelled data	The machine is trained on unlabelled data without any guidance	An agent interacts with its environment by producing few actions and discovers errors or rewards.
Type of problems	Regression and Classification	Association and Clustering	Reward based
Type of data	Labelled data	Unlabelled data	No pre-defined data
Training	External supervision	No supervisio	No supervision
Approach	Map labelled input to known output	Understanding the patterns and discovering the outputs	Follow trial and error method
Popular algorithm	Linear Regression, Logistic Regression, KNN, etc	K-means, C-means, etc	Q-learning, SARSA, etc

VI.REFERENCES

[1]. Introduction to Machine Learning – Third Edition – EthemAlpaydin

[2]. Machine Learning – Indian Edition – Tom M. Mitchell

[3]. Machine Learning with R – Second Edition – Brett Lantz

[4]. scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation

[5]. Stochastic Gradient Descent Algorithm With Python and NumPy – Real Python

[6]. Stochastic gradient descent - Wikipedia

[7]. Machine Learning Regression Explained - Seldon

[8]. What is Quantile Regression? | Introduction to Quantile Regression (mygreatlearning.com)

[9]. apriori-associative-learning-iq.opengenus

[10].unsupervised-learning-learn. g2

[11].computerscienceunsupervisedlearning-sciencedirect

[12].introductionclusteringandunsupervisedlearning-hub.packtpub

[13].DebomitDey:affinitypropagationinml, tofindthenumberofclusters.(May, 2019)

[14].Gilbert Tanner,ml-explained,mean-shift-explained.

[15].tutorialspoint:machinelearningwithpython-clusteringalgorithmsmeanshiftalgorithm

[16].Ahmad Hammoude, A Concise Introduction to Reinforcement Learning. (Feb 2018), DOI: 10.13140/RG.2.2.31027.53285.

[17].Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction, 2nd ed.

[18].Daniel Johnson , Reinforcement Learning: What is, Algorithms, Types & Examples. (Mar 2022)

[19].Ziad SALLOUM, Summary of Tabular Methods in Reinforcement Learning. (Dec 2018).

[20].Russell, Stuart J.; Norvig, Peter (2010). Artificial intelligence : a modern approach (Third ed.). Upper Saddle River, New Jersey. pp. 830, 831. ISBN 978-0-13-604259-4.

[21].Lee, Daeyeol; Seo, Hyojung; Jung, Min Whan (21 July 2012). “Neural Basis of Reinforcement Learning and Decision Making". Annual Review of Neuroscience. 35 (1):287– 30. doi:10.1146/annurev-neuro-062111- 150512. PMC 3490621. PMID 22462543

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Space Robotics - Guardians of Galaxy

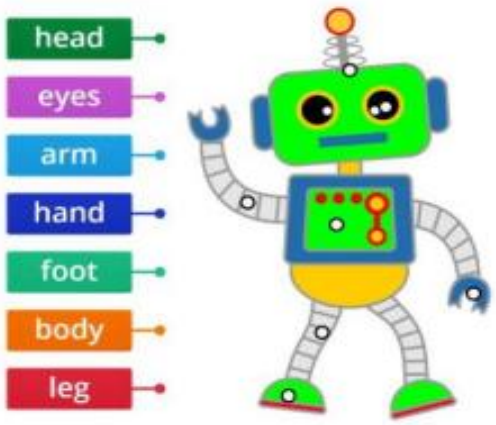
Aditi Vyas, Arpita Gado, Chidanandita Mohanty, Deepika Neelappa Gouda
Department of Computer Science & Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

CIMON, short for Crew Interactive Mobile Companion, is an AI based floating robot developed to reduce the stress of astronauts and make them feel a little less lonely. But CIMON turned out to be much more than just a support robot that smiled and replied to the astronauts when spoken to and a guide but a microgravity type robot that could perform the basic servicing around the spacecraft.
Keywords—CIMON, Floating robot, AI, Microgravity.

I. INTRODUCTION

Astronaut Scott Parazynski of STS-one hundred twenty performed a 7-hour, 19-minute spacewalk to repair (basically sew) a broken sun panel which facilitates the delivery of electricity to the International Space Station. NASA took into consideration the spacewalk risky with capability threat of electrical shock. The distance community, any unmanned spacecraft may be known as a robot spacecraft.



A. Why is it important

Sending a robot to an area is likewise tons inexpensive than sending a human. They can live to tell the tale in space for decades and may be ignored there—no need for a go back trip! Plus, robots can do masses of factors that people can't. Some can face up to harsh conditions, like severe temperatures or excessive stages of radiation.

CIMON was developed by the European aerospace company Airbus on behalf of the German space agency, which is known by its German acronym, DLR. The robot's AI is IBM's famous Watson system. CIMON is roughly spherical and weighs 11 lbs.

B. Types

Remotely operated vehicles (ROV) and faraway manipulator system (RMS) are the 2 principal varieties of area robots. A standard ROV may be a rover transferring over the terrain upon landing, a lander operated from a desk bound factor and is in touch with an extra-terrestrial simple or an unmanned spacecraft. Besides being utilized by area researchers for terrain exploration in the area, ROVs

are utilized by bomb squads to hit upon doubtlessly dangerous materials, in nuclear centers and subsea.

C. Examples

According to CNET, a few robots presently exploring the area encompass Dextre, Voyager 1 and 2, Hubble Space Telescope, Cassini, Robonaut 2, Rosetta, Dawn, Mars Express, Curiosity, Opportunity, 2001 Mars Odyssey, Advanced Composition Explorer, Hayabusa 2, Juno, Mars Orbiter Mission, New Horizons and greater to come.

II. METHOD AND MATERIAL

A. About

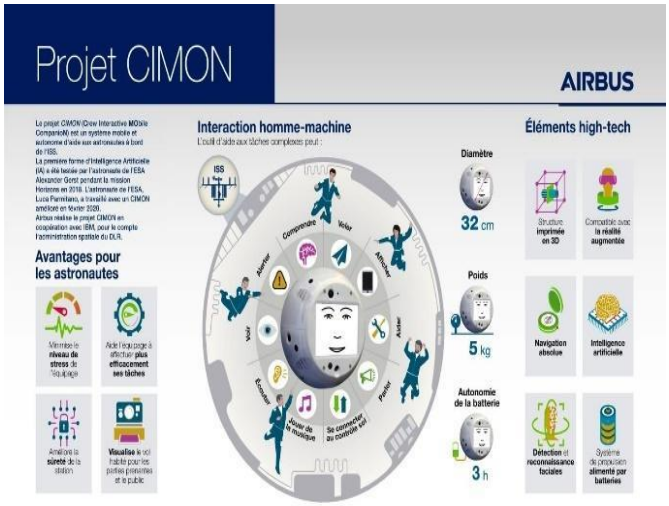
A picture of Cimon floating with astronaut Alexander Gerst.

The outer space is a vast mystery and man has forever wondered what lies beyond its planet Earth. Organizing manned missions was one way of decoding this puzzle. A manned mission normally had five astronauts on a tight schedule; with so many tasks to complete in a very short time, astronauts move from one task to another very quickly which eventually takes its toll on their efficiency. Airbus reached out to IBM and the German Aerospace Centre to build an innovative space assistant that will help reduce the pressure on the crew. Thus was born CIMON, an AI based floating robot. CIMON was named after Professor Simon Wright, the robotic assistant also called "The flying brain" from the Japanese science fiction series "Captain Future".

B. Features

CIMON was built with a total of 12 motors which helps it to move and rotate in any direction. Due to the help of these motors, CIMON can nod its head and also follow the astronaut on command. An LCD screen is placed to display the face and any other visual aids. It is colored in white. Cimon uses 2 front cameras as its eyes and one additional for face recognition and 2 at the sides for the video documentation. It uses microphones to detect sound and a directional microphone for voice recognition.

CIMON can be woken up with a wakeup call post in which the robot has the ability to help astronauts with the procedures or entertainment. CIMON can play an astronaut's favorite music and walk him step by step through a procedure with visual aid for help. It also helps in displaying the parts where they want to search making it easier. They can even search for an object in the shuttle or even attach an arm to perform different operations. Along with this, CIMON also has the capability to read human emotions using facial recognition technology helping it interact with astronauts. Cimon uses AI technology by IBM Watson Assistant which is used worldwide by many IBM clients. Through CIMON engineers want to test the extent of current capabilities of artificial intelligence, big data and data mining.



The Cimon is in the shape of a sphere. The description of CIMON is as follows in the table given below:

Diameter	32 centimeters
Weight	5 kilograms
Fans used	14
Microphones	7
Battery life	2 hours
Material used	Plastic & steel

Models

The CIMON consists of 2 models:

	CIMON1	CIMON2
Left	June 28th 2018 (via SpaceX Falcon9 rocket to the ISS)	December 5th 2019 (via SpaceX CRS-19 Mission)
Returned	August 27th 2019	-

CIMON 1 was updated to CIMON2 which has some additional features like autonomous flight capabilities, voice-controlled navigation and it can also understand and carry out various tasks. Another advantage of CIMON2 is that it has 30% more battery life than that of CIMON1.

III. RESULT AND DISCUSSION

We are currently living in the era of automation where autonomous vehicles and robots are taking over. These robots perform tasks which are dangerous for the humans to perform. CIMON is not currently equipped with self-learning capabilities and needs active human instructions which leave scope for improvement for the future space robotics' projects.

Types of robots simialr to CIMON

	Robonaut	Joint DARPA_NASA project designed to create a humanoid. It uses telepresence and robotic autonomy. The robot's hands have fourteen degrees of freedom and use touch sensors at the tip of the fingers. Currently 2 working robonauts, R1 and R2. R2 has flown to space for the STS-133 mission.
	Dextre	Two armed robots or telemanipulators on the ISS. It is helpful in spacewalks and also repairing objects on the surface of space stations.
	Rassor	Lunar robot that will excavate the soil with the Drum excavator. Either side there are arms mounted to climb over obstacles.
	Spidernaut	Arachnid inspired the Extra Vehicular Robot (EVR) designed by NASA for construction, maintenance and repair. Its legs can move at three points, one rotary joint at the hip and two more joints are linear actuated.

Main purpose of the robot during the first space mission in 2018, CIMON was only used for a specific range, but during the midterm, scientist have been planning to increase the range and trying to get CIMON for long spaceflight mission. This is because

scientist believed that interaction between a human and the assistance system is equipped with emotional intelligence. Main purpose was to make the astronauts more relaxed, since it's difficult to keep checking the manual about the correct location of the point that needs to be repaired. CIMON has the capability to save the map in the system. The features of the robot like being able to read human emotions and being able to interact with them in a way is good psychologically. There have been different robots made but mostly have legs or wheels. Being able to send a robot to space which can actually interact and has a capability to do many functions is a big advantage to save time and life. IV.

IV. CONCLUSION

In conclusion, scientists hope to use the CIMON robot for more kinds of activities like sending it to Mars and also being able to study human emotions and interactions more easily. Eisenberg once said that the service robots acting as a flight attend is a benefit for long spaceflight missions. It's impressive how CIMON, a small smart robot which is able to read human emotions and help astronauts in space missions, is created on a 3D printer and built with commercial available, off-the-shelf sensors and space-grade connector and cabling technologies in a span of less than two years. CIMON is the space robotics' first step into the world of AI powered space robots and is yet to see some major advancements in it. CIMON has proved its importance as a companion and a service robot in terms of efficiency. With advancements in artificial intelligence and machine learning algorithms we can definitely say, "CIMON says, leap into the future!"

V. ACKNOWLEDGMENT

The satisfaction and euphoria that accompany the successful completion of any task would be impossible without the mention who made it possible, whose constant guidance and encouragement crowned the efforts with success. I have great pleasure in expressing gratitude to Dr. Mohan Manghani, Chairman, New Horizon Educational Institutions, for providing necessary infrastructure and creating good environment. I take this opportunity ti express my profound gratitude to Dr.Manjunatha, Principal, New Horizon College of Engineering, for his constant support and encouragement. I would like to thank Dr. B. Rajalaskshmi, Professor and HOD, Department of Computer Science and Engineering for her constant support. I also express my gratitude to Dr. Meenakshi Sundaram B., Department of Computer Science and Engineering, my reviewer, for constantly monitoring the development of the research paper. His valuable suggestions were the motivating factors in my work.

VI.REFERENCES

[1]. Cimon (robot) – Wikipedia
[2]. IBM CIMON AI in space | IBM
[3]. https://www.discovermagazine.com/technology/meet-cimon-2-a-newand-improved-ai-robotastronaut?msclkid=aae223bdbb3f11eca874aa3b8df52b16
[4]. https://www.airbus.com/en/newsroom/press-releases/2020-04-cimon2-makes-its-successful-debut-on-the-iss
[5]. https://www.cnn.com/style/article/anicka-yi-tate-modern-turbinehall/index.html
[6]. List of NASA robots - Wikipedia
[7]. https://www.dlr.de/content/en/articles/news/2018/1/20180302_cimonthe-intelligent-astronaut-assistant_26307.html

[8]. https://www.youtube.com/watch?v=3_2Jy1Ur0js
[9] https://www.digitaltrends.com/cool-tech/new-improved-cimon-2- robot-iss/
[9]. https://www.abc.net.au/news/2019-12-06/cimon-emotion-seekingrobot-built-to-keep-astronauts-company/11773068
[10].https://longhorizon.org/trey/papers/mora19_guest_science_iac.pdf
[11].https://www.youtube.com/watch?v=FW5vYrseyk

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Congestion Control Using Network Based Protocol

Harshitha V Reddy[1], Manikandakumar M [1], S.J. Subhashini[1], Soja Rani[1]
[1]Department of Computer Science and Engineering New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The term "network congestion" refers to when the number of packets entering exceeds the number of packets leaving the network. Because of the increased use of multimedia applications, network congestion is increasing. There are a few rudimentary scheduling systems for controlling network congestion that only function at the end-to-end level, making them unable to avoid congestion collapse and unfairness caused by applications that are insensitive to network congestion. As a result, traffic flows became unresponsive and misbehaving. This method demonstrates the aggregation and connection admission control procedures in the NBP architecture using the network border protocol framework. The NBP framework offers an appropriate protocol architecture that requires the transmission of feedback between routers at network borders in order to identify and restrict unresponsive traffic flows before they join the network, reducing network congestion. Similar strategies have been integrated and improved them to provide a more flexible approach for constructing a congestion management scheme in a network utilising network border protocol.

Keywords: Internet, Network Border Protocol, Congestion Control, and Congestion Collapse.

I. INTRODUCTION

Parallel to the rise of the internet, there is a tremendous increase in multimedia consumption nowadays. TCP is a secure protocol that is mostly used for file transfers. This is seldom the case when transferring interactive video and music, when dependability is a disadvantage rather than a virtue. Because most network resources send undelivered packets, this is known as a congestion collapse.

Introduce and analyse the Network Border Protocol, a novel Internet traffic control method. If packets enter the network quicker than they leave, the network will buffer or, worse, reject the flow's packets. More packets are arriving than the network can handle. NBP assures that packets do not enter the network at a faster pace than they can exit. Because of the unresponsive flow, congestion from undelivered packets is reduced, whereas undelivered packets would ordinarily never reach the network. The first purpose of NBP is to prevent congestion from undeliverable packets causing congestion collapse.

UDP is classified as an unresponsive protocol since it lacks a way to detect or control congestion. The network may get congested if the limited capacity becomes a bottleneck, hence UDP control the transmission rate. It nearly exhausts the link's capacity. Congestion causes the selfclocking TCP to slow, lowering the good put to zero. The number of TCP flows into the internet is increasing, and the

congestion management mechanism, which is an inherent feature of the transport protocol, ensures network stability. The situation changes in the introduction of UDP, making the existence of several transport protocols almost impossible and causing congestion to rise. All data flows are monitored in NBP, and the transmission rate is changed based on a traffic shaper at the edge router. The DCCP, a combination of UDP and TCP with a congestion control feature, is the next option. Another option is still being explored and tried, which presents the question of an appropriate congestion control strategy when socially responsible and socially acceptable irresponsible protocols must coexist on the internet, as it does now.

II. LITREATURE

Dr. Chandrashekhar UPPIN [1] introduced the Congestion Free Router, a groundbreaking congestion-avoidance strategy (CFR). The CFR is the process of exchanging answers between routers at a network's perimeter in order to identify and restrict passive traffic flow before it enters the network, hence lowering network congestion.

Maxim Claeys [2] the congestion avoidance phase of the TCP New Reno congestion control mechanism was used to develop suggested deadline-aware congestion control strategy services. The congestion window can be dynamically altered to decrease the amount of deadline-missing flows by inserting deadline information in the transport layer. The suggested method only necessitates modifications on the sender's side and is network-invisible.

Suparna Banerjee [3] have published a comparison of many MANET routing protocol's performance. Because MANET nodes are so low powered, it's critical to avoid a large number of re-transmissions caused by network outages or congestion. Congestion control methods must be used optimally in order for the network and routing protocols to be reliable. Also Different works on various MANET routing protocols in the lack or presence of the TCP congestion management algorithm were investigated.

Michael C. Emmanuel [4] In contrast to prior congestion solutions, which were mostly end-to-end or host-to-host, the suggested network border patrol structure, which employs a rate control mechanism at the network's edge, has shown to be successful. The system avoids congestion by combining rate regulating and control at the in-router with per-flow rate monitoring at the out-router to guarantee packets do not enter the network faster than they leave.

Mohsen YaghoubiSuraki [5] because mobile networks are multi-hop networks, a high data transfer rate generates congestion. The method has three functions: detection, notification, and transmission rate modification. Congestion, buffer conditions, and the number of times the buffer fills are all monitored. The congestion level is determined by the fuzzy controller in the notification stage, and upstream nodes are notified. Finally, they change their transmission rate based on the congestion level of each node.

Yefa Mai [6] CC-AODV, an enhanced AODV routing algorithm, was presented. Simulations have been run, and the results of the AODV and CCAODV have been compared. When the network contains more nodes, the simulation results show that CC-AODV has a longer end-to-end latency than AODV. CC-AODV, on the other hand, beats AODV in terms of throughput, packet loss, and packet deliver ratio. The congestion counter is demonstrated to help minimise network "busy" nodes by increasing network throughput. Finally, the routing table's congestion counter adds to the overhead. Implementing a congestion counter in the routing table will be critical to achieving numerous routing pathways while

improving wireless performance. The preset counter threshold module will be optimised to improve the CCAODV implementation.

Kang Chen [7] The rationales underlying existing key correlated MPTCP CC algorithms were first examined in this study from the standpoint of subflow's capacity to compete for bandwidth. Three shortcomings of existing correlated algorithms are studied and validated based on the findings. This indicates that MPTCP requires decoupled CC. To that purpose, this paper presents a theory that decouples the CC of parallel subflow's nearly completely while keeping the advantages of correlated ones. CADIA's performance is evaluated through a series of studies. In the wireless context, CADIA outperforms correlated CC algorithms, according to the findings.

Takahiro Nogiwa [8] to accommodate appropriate bandwidths across high delay bandwidth networks, several high-speed congestion management techniques like as CUBIC and Compound TCP are utilised. The default congestion control methods in Linux and Microsoft Windows are CUBIC and Compound TCP. When CUBIC and Compound TCP traffic flows share a bottleneck link, however, Compound TCP's performance drops dramatically. To overcome this problem, locate them at the bottleneck connection and reduce CUBIC flow throughput. In this study, a convolutional neural network-based identification approach for congestion control methods using picture recognition.

Y. Narasimha Reddy [9] TCP-friendly operations may be performed by TFRC, and it can be balanced to fulfil various QoS requirements. By limiting bandwidth and transmission rate, TFRC is able to combat high traffic speeds and transmission delays. Propose an integrated TFRC that uses the WFQ technique to manage traffic congestion in high-traffic areas in order to improve TFRC performance. In comparison to TRFC, experimental examination of proposals reveals immediate processing with less packet loss and transmission latency. Even with TFRC's lengthy RTT difficulty and packet loss, this technique decreases transmission rate suppression.

Prashant Kharat [10] With every ACK frame reception, the window update frame adjusts the window size, reducing one protocol transition state. With each ACK receipt, the recommended modification window size changes, resulting in a smooth variation in congestion window size. This smooth fluctuation in the congestion window manages network traffic and aids in congestion control. The only responsibility of the congestion management mechanism is to keep streaming data running smoothly. When there is no congestion, however, a step-by-step approach for updating the delivery window is used.

Samuel Jero [11] suggested an automated approach to detecting attacks against TCP congestion control implementations that combines the scope of implementation-agnostic interference with the precision of runtime analysis. It employs a model-guided approach to generate abstract attack strategies, utilising a state an attacker might use a machine model of TCP congestion management to uncover susceptible state machine pathways exploit to increase or decrease the throughput of a connection to his advantage to automatically discover manipulation attacks on congestion control without requiring the user to provide any vulnerable lines of code and without being dependent on the specific implementation, language, or implementation. Protocol facing is a well-known technique in which packet contents are randomly created and injected into the network, or are randomly altered in transit.

Weinan Zhang [12] provided a technique for selecting a secure channel under dynamic congestion management. The current amount of congestion is determined by calculating the average queue length at

the nodes. If the network anticipates future congestion, it will choose a new path in which all nodes have been certified as trustworthy nodes, producing session keys in the TCP three-way handshake to prevent denial of service attacks. The average queue length was proposed as a novel secure congestion control approach for predicting the degree of network congestion.

Lal Pratap Verma [13] compared to previous TCP variations, introduces a unique congestion control strategy to tackle the significant round trip delay and give improved bandwidth usage. RTT variation is used as an indication of congestion delay-based TCP variations, which employ RTT as a congestion detection component. The suggested TCP instantly adjusts the transmission rate based on the network traffic conditions. The simulation results demonstrate that the suggested TCP minimises packet losses and increases bandwidth usage in variable RTT and background traffic scenarios. With regular TCP Reno, it also demonstrates improved inter-protocol fairness.

Mahadev A. Gawas [14] in MANETs, a new cross-layer routing protocol dubbed congestion-adaptive and delay-sensitive Multirate (CADM) was suggested. The CADM protocol takes use of the network, MAC, and physical layers' cross-layer interaction. In MANETs, CADM uses the connection between data rate, congestion metric, and MAC latency to improve network efficiency. In Multirate MANETs, the protocol identifies different node-disjoint routes and supports appropriate data rates between the connections depending on the projected delay. The proposed CADM protocol finds a path through less crowded nodes and actively tackles congestion when it happens. The CADM protocol's performance is thoroughly evaluated through simulation, highlighting the benefits of the cross-layer method.

OluwaseyiAkinlade [15] proposed a novel way to adjusting transmission power that is based on the network's vehicle density. The goal is to minimise network channel congestion and increase overall network performance. According to simulation data, this strategy can improve performance in terms of packet loss and inter-packet latency. In this research, a novel method for controlling channel congestion in VANETs by adjusting transmission power, according on the network's vehicle density. Based on the density of cars on the roadways, the vehicles will broadcast the packets at appropriate transmission strengths.

Kaoutar BAZI [16] studied the behaviour of different congestion control strategies in crowded wireless mesh networks to find their benefits and drawbacks. The well-known network simulator ns2 is employed in the simulation. TCP Tahoe, TCP Reno, TCP New Reno, and Sack are loss-based protocols that are beneficial for latency-sensitive flows, whereas TCP Vegas, which is delayed-based, is recommended for applications that do not suffer from information loss, but have fairness issues when sharing a bottleneck with competing flows. Several congestion control strategies have been proposed to alleviate congestion.

Yi-Cheng Tsai [17] TCP version that is better suited for data centres and delivers reliable transmission with high throughput and low latency are offered. Reviewing several TCP congestion control algorithms and selecting BBR as the foundation for the proposed SRTT congestion control mechanism. The explanation behind BBR and how it can't achieve near-zero queueing as intended. Proposed an SRTT congestion control strategy that prioritised maintaining a steady RTT before attempting to enhance throughput. In a data centre network environment, SRTT was shown to have a smaller queue size and greater throughput than BRR and TCDCP.

Jin Zhu [18] developed an average queue-length-difference–based congestion detection system in which the congestion signal is selected as an

exponentially weighted moving average queue-length difference, with the theoretically optimal forgetting factor derived. The state equation is obtained from the fluid model by defining the queue-length difference at the state. The inflow traffic in the state equation is a discrete-time martingale that, using the martingale representation theorem, may be turned into a Wiener process. With the presence of unreliable transmission in the TCP network, an average queue-length-difference–based congestion detection technique.

Caitlin Facchina [19] suggested a novel distributed congestion control technique that manipulates transmission power using a density estimation depending on the vehicle's speed. The findings show that the suggested strategy reduces packet loss and improves the relevancy of receiving messages in VANETs, where vehicles vary their transmission power based on their speed. The notion is that the quicker a vehicle travels, the less crowded the network becomes. With greater space between cars, larger transmission powers may be employed, but a low-speed traffic network may have considerable congestion at the same power.

Prashant Kharat [20] The proposed Modified QUIC (ModQUIC) protocol's effectiveness in terms of congestion control has been evaluated using India's fast increasing Internet service provider, the Reliance Jio 4G network (JioFi), which has quickly taken 17 percent of the market share. The performance of the ModQUIC protocol in the JioFi network was evaluated using congestion management techniques CUBIC and Bottleneck Bandwidth Round-trip-propagation-time (BBR). The experiment is carried out on a testbed built with JioFi and a RaspberryPi-3 wireless router, as well as a network emulator called Netem. Test video clips kept Google Drive are used to verify the ModQUIC's performance. Throughput and Retransmission Ratio measurements are used to evaluate performance under packet loss and reordering scenarios (RTR).

III. PROBLEM DEFINATION

As a result of its adamant adherence to end wise crowding management, the modern Web suffers from two pathologies: congestion induced by undelivered packets and inequitable bandwidth distribution among competing traffic flows. When undeliverable packets eat up bandwidth, congestion rises. For a variety of reasons, including the prominence of claims that do not respond to traffic appropriately, one-sided bandwidth allocation networks are becoming more widespread on the internet. In the fight against slow-to-respond apps, adaptive apps that drop their transmission rates quickly in reaction to congestion are likely to be assigned disproportionately modest bandwidth allotment. Unfairness can be brought discussing internet protocols in a multitude of ways. The TCP algorithm, for example, automatically gives a bandwidth to each TCP flow that is inversely proportional to its round-trip length. As a result, TCP connections with short round-trip lengths may receive excessively large network bandwidth allocations compared to connections with longer round-trip periods. The impact of expanding streaming video traffic on traditional data traffic is causing significant anxiety among Internet users. Because streaming media traffic is hypersensitive to network congestion, congestion collapse and the inequitable bandwidth allocation might be exacerbated.

IV. RESULTS AND IMPLEMENTATION

- **Proposed System**

CFR compares the rate at which packets from each application flow enter and exit the network. When

packets enter the network quicker than they leave, the network acts as a buffer or, even worse, discards the entire flow. To put it another way, the network is being bombarded with more packets than it can process. CFR avoids this problem by monitoring the network perimeters, ensuring that packets do not enter at a faster pace than they can exit. Because unresponsive traffic would otherwise never enter the network, patrolling minimizes congestion collapse caused by undelivered packets. And created a single protocol, Network Border Protocol, to eliminate the utilisation of several methods such as Core Stateless Fair Queuing, Round trip time, and packet data prevention (NBP)

- **System Architecture**

End-to-end congestion management strategies may be solved by moving congestion control from host systems to the network, which they term network border patrol (NBP). NBP measures the frequency with which packets enter and depart the network using entry and exit routers (in routers and out routers, respectively). A Rate Control Algorithm (RCA) and a Time Sliding Windows (TSW) Algorithm were used in the previous system. The rate control algorithm keeps track of and manages the entry and exit flows, while the Time Sliding Window method guarantees that packets are delivered in the correct order. NBP assumes that packets from applications enter and exit the network at different speeds.

Figure 4.1: System Architecture

When packets leave slower than they join the network, NBP believes the network is buffering. This indicates that the network is experiencing a surge of packets. NBP enters the network by "patrolling the network's borders," monitoring the flow rate of the main-input router's and output, and ensuring that entry and leave rates are balanced. Unresponsive flows are not permitted into the network, which prevents congestion collapse. However, in order for an edge router to know the pace at which packets leave the network and communicate feedback with other edge routers, NBP adds an additional communication cost.

- **Implementation**

Congestion Free Router Internet traffic control protocol to alleviate the ills of congestion collapse (CFR). CFR's core premise is to compare a network's boundaries, as well as the rates at which packets from each application flow enter and exit the network. If packets enter the network quicker than they leave, the network is probably buffering or, worse, discarding the flow's packets. In other words, the network is being bombarded with more packets than it can handle. CFR prevents this scenario by monitoring the network's boundaries, ensuring that packets from each flow do not enter at a faster pace than they depart. Because unresponsive flow's otherwise undeliverable packets never enter the network in the first place, this patrolling eliminates congestion collapse due to undelivered packets.

The CFR is capable of averting congestion collapse and increasing bandwidth distribution fairness, however these benefits are not free. CFR overcomes these difficulties at the cost of increased network complexity, because CFR requires routers at the network's edge to monitor and manage individual flow rates. CFR adds to the communication overhead since an edge outer has to know the rate at which its data is being sent as packets leave the network, the

edge router must communicate with other edge routers. Some of the current efforts to resolving traffic congestion have failed. CFR complexity is limited to edge routers; routers at the network's core are not involved in congestion prevention. Furthermore, end systems are completely unaware that CFR has been implemented in the network, hence no modifications to transport protocols are required at the end systems.

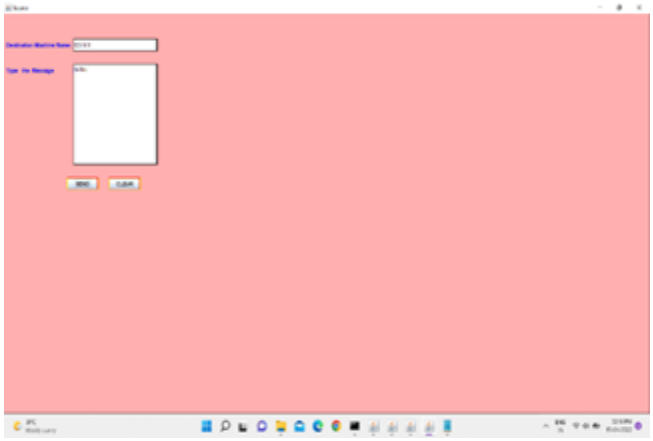


Figure 4.2: Source Page

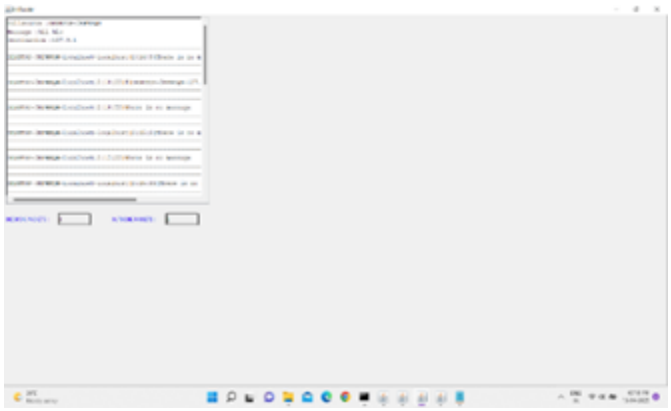


Figure 4.3: INRouter Page

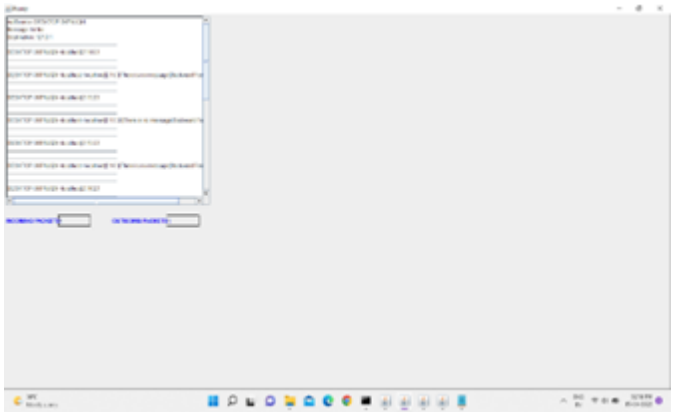


Figure 4.4: Router Page

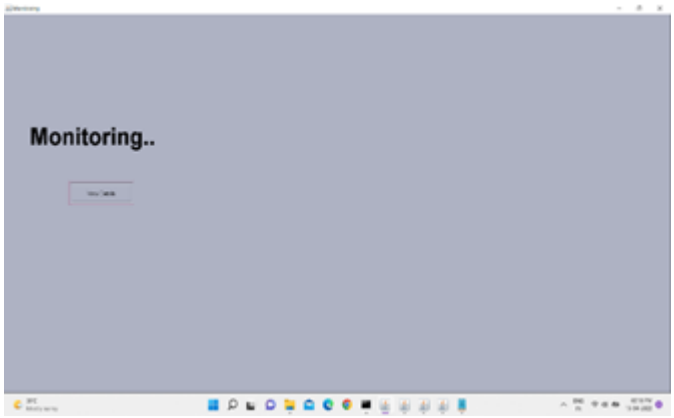


Figure 4.5: Monitoring Page

V. CONCLUSION

Retransmission of lost or dropped packets, as well as increasing the queue size, are not accessible to avoid congestion. There is also a requirement for speedier data transfer without packet loss during transmission. By meeting all of the user's demands, NBP is able to avoid congestion collapse. NBP ensures that the bandwidth available to the systems is used effectively. As a result of the research, employing NBP to avoid packet drop is a superior strategy for preventing packet loss in the network.

VI.REFERENCES

[1]. Dr. Chandrashekhar UPPIN HOD, Department of Computer Science, Faculty of Computing and Applied Sciences. Baze University, Jabi Abuja (Nigeria)

[2]. Maxim Claeys, IDLab, Department of Mathematics and Computer Science, University of Antwerp–imec, Antwerp, Belgium. Niels Bouten, IDLab, Department of Mathematics and Computer Science, University of Antwerp–imec, Antwerp, Belgium and IDLab, Department of Information Technology, Ghent University–imec, Ghent, Belgium. Danny De Vleeschauwer, Bell Labs, Nokia, Antwerp, Belgium. Koen De

Schepper, Bell Labs, Nokia, Antwerp, Belgium. Werner Van Leekwijck, IDLab, Department of Mathematics and Computer Science, University of Antwerp–imec, Antwerp, Belgium. Steven Latré, IDLab, Department of Mathematics and Computer Science, University of Antwerp–imec, Antwerp, Belgium. Filip De Turck, IDLab, Department of Information Technology, Ghent University–imec, Ghent, Belgium

[3]. Suparna Banerjee, Arindam Ghosh, Raja Karmaka and Sourav Mahapatra. Department of Information Technology Techno International Newtown Kolkata

[4]. Michael C. Emmanuel Department of Computer Science University of Port Harcourt Nigeria L. N. Onyejegbu Department of Computer Science University of Port Harcourt Nigeri

[5]. Mohsen YaghoubiSuraki, AbolfazlToroghiHaghighat, Majid Gholipour M YaghoubiSuraki, A. ToroghiHaghighat, M. Gholipour. Department of IT and Computer Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

[6]. Yefa Mai, Fernando Molina Rodriguez and Dr. Nan Wang. Department of Electrical and Computer Engineering Calfornia State University, Fresno Fresno, USA

[7]. Kang Chen and MijanurRahaman Palash. Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL 62901

[8]. Takahiro Nogiwa and Kouji Hirata. Graduate School of Science and Engineering, Kansai University, Osaka, Japan. Faculty of Engineering Science, Kansai University, Osaka, Japan

[9]. Y. Narasimha Reddy Rayalaseema University, Kurnool, A. P., India E-mail: narasimhareddy.phd@gmail.com P V S. Srinivas Sreenidhi Institute of Technology and Science, Hyderabad, India 08 June 2018

[10]. Prashant Kharat and Muralidhar Kulkarni. Department of Electronics and Communication, National Institute of Technology Karnataka, Surathkal, Mangalore, India. “Modified QUIC protocol for improved network performance and comparison with QUIC and TCP”

[11]. Samuel Jero, Purdue University. Endadul Hoque, Florida International University. David Choffnes, Alan Mislove and Cristina Nita-Rotaru, Northeastern University

[12]. Weinan Zhang and Li Du, Computer Science and Engineering, Northeastern University, Shenyang, China. Li Lin, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA

[13]. Lal Pratap Verma, Moradabad Institute of Technology, Moradabad, Uttar Pradesh, India. Indradeep Verma, IIMT College of Engineering, Greater Noida, Uttar Pradesh, India. Mahesh Kumar,Jaypee University of Engineering and Technology, Guna, Madhya Pradesh, India

[14]. Mahadev A. Gawas, School of Computer Science and Engineering, VIT, Vellore, India. Lucy J. Gudino, Department of Computer Science and Information Science, BITS Pilani K. K. Birla Goa Campus, Goa, India. K. R Anupama, Department of Electrical and Electronics, BITS Pilani K. K. Birla Goa Campus, Sancoale, India

[15]. OluwaseyiAkinlade, Ikjot Saini, Xiaofeng Liu, ArunitaJaekel Department of Computer Science University of Windsor Windsor, Canada

[16]. Kaoutar BAZI and Bouchaib NASSEREDDINE. Applied Mathematics and Computer Science laboratory FST, Hassan 1st University Settat, Morocco.

[17]. Yi-Cheng Tsai, Chiayi, Ting-Chao Hou. Mao-Ching Chiu. National Chung Cheng University 168 University Road Min-Hsiung, Chiayi, Taiwan
[18]. Jin Zhu and Tong Luo, Department of Automation, University of Science and Technology of China, Hefei, China. Lin Yang, Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. WanqingXie, 3School of Information Science and Technology, University of Science and Technology of China, Hefei, China. Geir E. Dullerud, Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[19]. Caitlin Facchina, School of Computer Science University of Windsor Windsor, Ontario, Canada. ArunitaJaekel School of Computer Science University of Windsor Windsor, Ontario, Canada
[20]. Prashant Kharat, Department of Information Technology, Walchand College of Engineering, Sangli, India. Muralidhar Kulkarni, Department of E&C Engineering, National Institute of Technology Karnataka, Surathkal, India. "Congestion control performance investigation of ModQUIC protocol using JioFi network: A case study"

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Analysis of the Various Ethics, Tools and Attacks in Ethical Hacking

Shriya B Krishnan, Ms. Uma N
Department of Computer Science and Engineering, Viswesvaraya Technological University, Bangalore, Karnataka, India

ABSTRACT

Our current age is also known as the 'Internet Age' and even though it has paved way for so many advancements in technology and has enhanced the world, it still has some shortcomings. An example for this can be security related issues and privacy concerns. Hacking also referred to as invasion of one's privacy be it personal or an infrastructure has definitely made the internet to become a bane of its' existence. This is where Ethical Hacking plays as role wherein through the perspective of a Hacker and using various exploitation methodologies one can be enlightened more on the basic concepts. And hence this paper has been undertaken to gain more insights on Ethical Hacking, the basic ethics of it, the various attacks such as: Malware, Phishing, DoS (Denial of Service) and the methodologies used in Ethical Hacking such as: Reconnaissance, Scanning and Enumeration, Gaining Access, Clearing Tracks. The main purpose for creating this Research paper is to garner more knowledge and understand from the perspective of a Hacker on how to protect our privacy and business infrastructure along with having a clear comprehension of the differences between Penetration Testing and Ethical Hacking and understating this through various styles of Penetration Testing as well.

Keywords: Ethical Hacking, Hacker, Malware, Phishing, Vulnerability

I. INTRODUCTION

"Ethical Hacking" these words together are quite contradictory to each another as the word 'Ethical' refers to something moral whereas the word 'Hacking' refers to invasion of privacy. In the present world where there's a boom in Computers as well as the Internet, there is definitely high chances of privacy and security being compromised. And this is where hacking and Ethical Hacking in general come into picture.

Ethical hacking refers to the process of using various techniques and principles of hackers to help various businesses and people to protect their infrastructure and information with their consent. When considering the ethics of Ethical Hacking, it is something that is always performed by trained professionals and is always performed with **consent.** In order to hack ethically we have 'Ethical Hackers' who are skilled professionals with immense knowledge in technology and who also have the ability to identify and exploit vulnerabilities in a system by complying the ethics and rules of a given organization or an individual. The main intention of Ethical Hacking is to search out the vulnerabilities from a view point of a hacker (who is someone who

intentionally intrudes upon others privacy and hacks/steals into their infrastructure for monetary/personal gain) and therefore secure the systems and provide overall risk management as well.[1]

Some of the important terminologies to be considered in hacking are as follows:

Threat: This refers to anything that has a potential to cause harm and the various threats that are available are System Threats, Network Threats, Application Threats, Cloud Threats, Malicious file Threats etc.

Vulnerability: Vulnerability refers to a weakness or a flaw in the system through which an attacker can find and exploit the system. Examples of vulnerabilities are: an updated OS, unencrypted protocols, default passwords.

Attack: This term refers to a method that is followed by a hacker/individual to intrude into a system and hack into sensitive information. [5]

II. VARIOUS ATTACKS IN ETHICAL HACKING

Attacks refer to the methodology to find vulnerabilities by exploiting a given system. The various attacks are as follows:

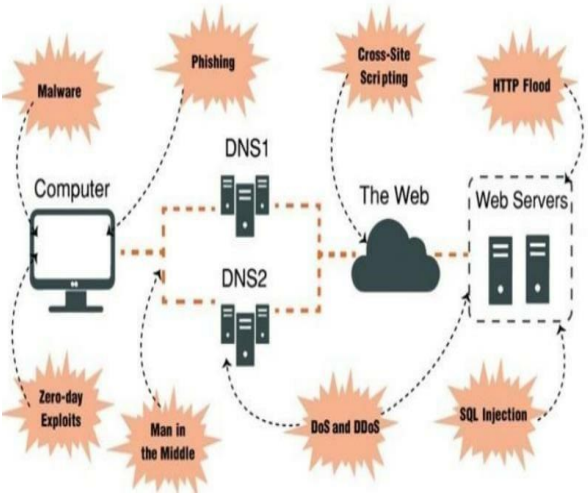


FIGURE 1: Types of Attacks

1. **Malware:** Malware is considered to be a malicious software that can enter into a target host and give the attacker full or limited control over the target. They are used to either damage or modify the information and functionalities of the target host and thus helping the attacker to access or destroy information. The various types of malware are:
   - **Virus**
   - **Trojans**
   - **Worms**
   - **Rootkits**
   - **Spyware and**
   - **Ransomware**
2. **Phishing:** Phishing is described to be the practice of sending fraudulent/deceiving information that appears to be very convincing and can thus lure in individuals that give out their details and fall prey to such inconveniences. Examples include: Fake emails from reputable organizations, asking users to give up their login credentials for social networking sites etc.
3. **Man in the Middle:** This type of an attack can also be considered as a personal attack wherein an individual places themselves between a user and an application to either eavesdrop and learn about their information or impersonate them. This is one of the reasons why it is advised to always be vary of people around and not to share passwords or passcode of anything.
4. **Cross-Site Scripting:** It is also referred to as 'XSS Attacks' as this kind of an attack takes place in web applications wherein the hacker sends/injects a client-side malicious script (also known as malicious payload) as an input into a legitimate web application, and thus cause various damages for the victim causing data leakage, hijacking a session etc.
5. **DoS (Denial of Service) and DDoS(Distributed Denial of Service) –** This is an attack wherein the

hacker creates a lot of traffic for the system such that it overloads the resources and the bandwidth that is available thus making the servers and the network unavailable for the users.

6. **SQL Injection:** SQL Injection refers to an attack where the attacker makes use of SQL and inserts code lines in order to allow the data to be displayed so that they can manipulate the victims' data and use it to their advantage. This attack is classified into two types, they are:
   1. **Error Based –** Here the attacker sends in malicious queries to a database which then results in errors and this gives useful information on the victim
   2. **Blind Based –** This type of SQL injection leads to cluelessness as to whether the web application is said to be vulnerable to the attack or not. This is further divided into two types:
      - **Boolean Based:** In this case only correct queries are displayed and
      - **Time Delay:** This depends upon conditions and setting a time delay on it. If the conditions reinforced are satisfied then the time delay is observed.[10][11]

III. VARIOUS METHODOLOGIES AND THE TOOLS USED IN ETHICAL HACKING

The various methodologies are nothing but the processes/phases through which Ethical Hacking is carried out.

FIGURE 2: Phases of Ethical Hacking

1. **Reconnaissance –** Reconnaissance is described as the set of processes and techniques that are used to discover confidential details about a target source. During this phase the hacker tries to gather as much information as they can. The tools deployed for this are:
   - **Google –** Gives basic information available regarding something.
   - **WhoisLookup –** This is a lookup tool that is used to check the availability of a domain name and all the information of the given host such as IP Address, ownership info etc.
   - **NSLookup –** This is a network utility program to gather information on internet servers and find all the information regarding it by querying the DNS (Domain Name System). [2]
2. **Scanning and Enumeration:** The second phase contains two terms where scanning refers to the process of searching for open doors which is nothing but discovering the in the services that are running on a port and enumeration refers to the process of garnering more information about a target machine and this can be achieved by connecting to it actively. [4]
   The tools used for this are:
   - **Ping –** This is used to check if the host computer is alive and can process any requests that are sent.
   - **Nmap –** Nmap is a shortform of "Network Mapper" which is a well-known free open-source hackers tool used for network discovery and security auditing.
   - **Nikto –** It is an open-source webserver scanner tool used for scanning and checking for vulnerabilities.
3. **Gaining Access:** Once the two phases are completed the hacker tries to gather enough information to gain all the access on the target.

An example for this is Password Cracking. The tools used here are:

- **John the Ripper –** Most commonly referred to as 'John' this is a popular penetration testing tool that is used to crack password and carry out dictionary attacks.
- **Wireshark –** It effectively captures packets in a real time network and displays data about packets travelling in human readable format.
- **Aircrack –** This uses the best algorithms to recover wireless passwords by capturing data packets.[6]

4. **Maintaining Access:** Once the hacker has gained the access, they can further employ more techniques to hack into nearby systems and exploit their vulnerabilities or continue to exploit the target system. The various tools used for this are:
   - **Metaspoilt Penetration Testing Software –** It is a Cyber Security framework that provides the individual with vital information with the known vulnerabilities and helps to formulate various strategies and methodologies for exploitation.
   - **Beast –** It is one of the examples of a Trojan Horse which is used to create backdoors also known as 'Remote Administration tools' i.e., 'RAT'.
   - **Cain & Abel –** This is a tool that is used for recovery of passwords by using methodologies such as network packet sniffing and also to crack password hashes
5. **Covering/Clearing Tracks:** The final phase is the process in which the attacker has to cover their tracks and destroy any evident evidence of their presence and this step usually ensues by erasing login credentials or any other error messages that might have been previously traced back. This can be achieved by using the following tools:
   - **Metaspoilt Penetration Testing Software –** It is a Cyber Security framework that provides the individual with vital information with the known vulnerabilities and helps to formulate various strategies and methodologies for exploitation.
   - **OSForensics –** This tool is used for deletion of the log and registry files. [4]

What is Penetration Testing and how is it different from Ethical Hacking?

Penetration Testing or Pen testing refers to a form of cyber security assessment that can be done ethically to identify and safely exploit the vulnerabilities in an organization and thus eliminate the vulnerabilities that are encountered.

Though both Penetration Testing and Ethical hacking are closely related and are often interchangeable, Ethical Hacking mainly focuses upon the various hacking methods and attacks and how they can be eliminated. This is done by thinking in the perspective of an actual hacker whereas Pen testing more or less just focuses upon discovering the vulnerabilities, the risks and target environment and thus successfully securing and taking control of the system. [7]

There are various types of pen testing but among them there are three main types based upon which pen testing occurs by the amount of information that is provided.[8]

The three types are:

1. **White Box Penetration Testing:** This testing method is often referred to as the crystal or oblique box pen testing which involves complete network sharing and information sharing with the given tester. This type of testing style is very useful for simulating an attack methodology on a

specific system by utilizing as many attack vectors as possible.[12]

2. **Black Box Penetration Testing:** In case of Black Box Pen Testing, no prior information is shared with the tester at all. In this scenario the tester applies the methodology of an unprivileged attacker/hacker that is from the initialization through exploitation process.
3. **Grey Box Penetration Testing:** Grey Box Pen testing is also called as 'translucent box test' as only a limited amount of information is shared with the tester. This can be in the form of login credentials or some basic information. This type of pen testing allows the tester to understand what level of access a user could gain and the potential damage that they may cause as well. This methodology is often favoured as it provides the balance, authenticity and also any time-consuming process. [3][9]

IV. CONCLUSION

We know that our current world is moving more and more towards the enhancement of technology and this has definitely led to the digitalization of a lot of processes. With this we definitely have the increased risk of privacy and security. This research paper has been written in order understand about Ethical Hacking and the different phases through which it takes place. The analysis of the different attacks and the tools used in the methodologies as well but most importantly this paper has helped us know how attacks from a hackers' perspective happen and easily susceptible systems are as well. The difference between Penetration Testing and Ethical hacking has also been discussed analytically with examples such as various styles of penetration testing with respect to information sharing.

V. REFERENCES

[1]. https://www.greycampus.com/opencampus/ethical-hacking/what-is-hacking
[2]. "Hacking for Dummies" a book by Kevin Beaver, CISSP (Information Security Consultant)
[3]. The Ethical Hack: A Framework for Business Value Penetration Testing – Book by James S. Tiller
[4]. Hacking: The Art of Exploitation – Book by Jon Erickson
[5]. The Basics of Hacking and Penetration Testing – Book by Patrick Engebretson
[6]. Hackers: Methods of Attack and Defense. Online. Discovery Communications.28Oct.2003
[7]. Ethical Hacking Techniques with Penetration Testing www.ijcsit.com/docs/Volume%205/vol5issue03/ijcsit2 0140503161.pdf(by KB Chowdappa)
[8]. Ajinkya A.Farsole, Amurta G. Kashikar and Apurva Zunzunwala , "Ethical Hacking " , International journal of Computer Applications (0975-8887), Vol. 1 No. 10, pp. 14-20, 2010
[9]. H.M David, "Three Different Shades of Ethical Hacking: Black, White and Gray," in GSEC Practical Assignment, Version 1.4b, Option 1, Feb 23, 2004
[10].System Security and EthicalHackingwww.ijreat.org/Papers%202013/Volume1/IJR EATV1I1018.pdf
[11].Hacking: Computer Hacking, Security Testing, Penetration Testing, and Basic Security – Book by John Slavio.
[12].https://www.eccouncil.org/what-is-penetration-testing

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Study on Self Driving Cars

Rekha M.S[1], Manjunatha B N[1], Charan Kumar Reddy K[2], Pavan Kumar Reddy K[2]

[1]Assistant Professor, Department of Computer Science and Engineering, R.L Jalappa Institute of Technology, Doddaballapur, Karnataka, India

[2]UG Student, Department of Computer Science and Engineering, R.L Jalappa Institute of Technology, Doddaballapur, Karnataka, India

ABSTRACT

There's some boom in Auto driving car attentiveness during last 10 years [1]. This is the most modern and cutting-edge skill in the car business, and it was successfully developed by TESLA despite having been discussed and worked on for a long time [13]. An Autonomous car recognizes its environment and can drive through traffic and other obstacles with little or no human involvement.

This sort of device has the potential to revolutionize mobility for those with disabilities and allow blind people to walk freely. This sort of vehicle improves the overall safety, security, and productivity of the motor transportation industry, and human errors may be removed while the drive is optimized.

We propose an autonomous car prototype that combines disparate technologies such as a road lane detection algorithm, a disparity map algorithm to determine the distance between the car and other vehicles, and anomaly detection using the Support Vector Machine classification algorithm, which performed well on our data set [7]. Autonomous vehicles are getting closer to becoming approved, but they aren't yet safe enough to be utilized in real life due to a lack of safety.

Keywords: Autonomous car, Cutting-edge technology, Prototype, Lane detection algorithm.

I. INTRODUCTION

Autonomous Cars, often known as Autonomous cars or driverless cars, are an exciting new technology. Vehicles that can drive themselves from one area to another without the intervention of a person are known as Autonomous Automobiles.

Self-driving cars have received both favourable and negative feedback on its operation and ethical usage after being debated by a number of public personalities, media sources, and academic experts [11].

In order to [2] improve traffic regulations, increase passenger safety, and simplify travel, programming will be utilized to collect statistics from the environment, routes, roads, signs, and other cars when performing a simple calculation.

This should be more efficient and beneficial than a human-driven vehicle. Meanwhile, as the general public becomes more aware of the countless incidents

caused by self-driving cars, ethical questions about their use have grown.

Figure 1: Technology in self-driving cars

The development of Autonomous automobiles may have generated hope in potential investors and technology enthusiasts, but many fail to consider the various limits that these vehicles may confront. Many of these Autonomous cars, for example, are built to operate under specified conditions and with precise instructions; but, how will they be able to make a valid judgement if these criteria are not fulfilled or if more than one solution for a job is available?

They are autonomous cars that perceive their environment using a variety of technologies such as radar, object recognition, and navigation systems. Amongst the most common causes of fatal accidents is driver error. The attention of drivers is diverted to mobile phones, gadgets, and audio systems. As a result, if the drivers are not paying attention, Autonomous cars equipped with self-parking and pre-safe technologies can take charge.

Though some argue that Autonomous vehicles are not as safe as human-driven cars, they are safer with fewer errors and faster reflexes since there is no human involvement, only the car's software, which decreases traffic and road rage and promotes safety. They may also have a variety of financial benefits. In the 1990s, the early attempts to build vehicles focused on constructing radio-equipped highways that could communicate with the cars that drove on them. These circuits were buried in the 400-foot road to create an opportunity for Autonomous vehicle manufacturers.

Despite the fact that these test drives yielded excellent results and that the cars can navigate on specific routes, the automobiles created to move using radio signals were not truly autonomous. Because they couldn't react to their environment on their own, they had to rely on commands and signals.

VaMoRs is an unmanned vehicle with two cameras, sensors, and an onboard computer that was created in 1987. This vehicle was capable of recognizing and reacting to other vehicles, traffic signs, and road conditions independently.

In 1988, a group of engineers led by Ernst Dickmanns unveiled the first self-driving Mercedes. This car covered 994 miles and reached a top speed of 112 miles per hour. Only 5% of the journey was completed without the use of human aid. The most well-known manufacturers of autonomous vehicles today are Google and Tesla Motors.

Google's self-driving cars rely on a combination of Google's street view maps, sensors, and cameras strategically positioned throughout the vehicles. The acronym "Self-Driving Coalition for Safer Streets" means "Self-Driving Coalition for Safer Streets.". All of these characteristics make self-driving vehicles safer than human-driven vehicles.

II. LITERATURE SURVEY

A lot of efforts have been made to bring some differences between conventional automobiles and self-driving cars. So far, it may be found in all the data acquired by examining various facets of this specific sector.

It appears that self-driving vehicles are far harmless than human driven cars, but more study is needed to establish this.

This is owing to advances in deep learning, where deep neural networks are trained to execute tasks that would normally need human interaction. CNNs are important in the field of Computer Vision because they employ models to find patterns and characteristics in pictures.

In 2009, Google launched its self-driving car project, which included employees who had been working on the technology for years. The Google automobile will be on the road for testing by 2012. With the passage of time, the car has evolved and is now equipped with several sensors, radars, lasers, a Global Positioning System (GPS), highly detailed maps, and a variety of other features to safely drive and navigate itself without the need for human intervention. The automobile can not only drive itself, but also park and drive on motorways. Cameras are utilized to locate and recognize things, which are subsequently analyzed by the car's computer [3]. Google unveiled a fresh idea for their self-driving automobile in May 2014, which had no steering wheel or pedals and unveiled a fully functioning prototype in December of that year that they planned to test in 2015.

These self-driving cars are safe as long as their sensors and software are up to date.

III. VARIOUS STAGES OF SELF DRIVING CARS

During 1925, a man called Francis Houdina developed a radio-controlled automobile that could be driven down the street with no one in the driver's seat, laying the framework for self-driving cars. General Motors debuted the first self-driving vehicle model at the 1939 World's Fair. It was a magnetized metallic point embedded in the road that pushed a radio-controlled electromagnetic field-directed electric vehicle. In 1958, this model became a reality.

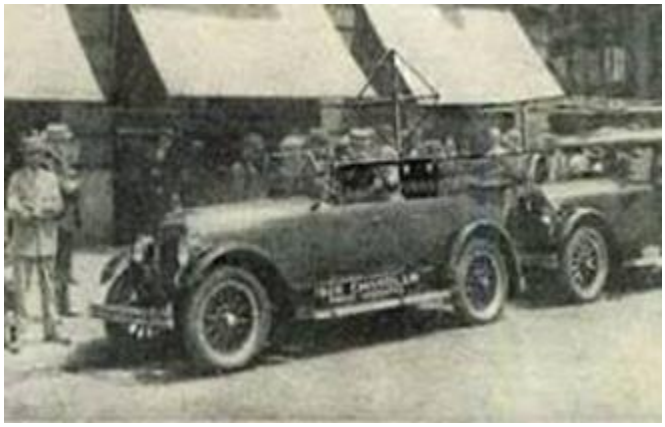


Figure 2: First radio-controlled car without driver

Norman Bel Geddes' self-driving vehicle for General Motors was finally completed in 1958, more than two decades after the concept was initially displayed at the New York World's Fair. It was steered by a technology that adjusted the electromagnetic fields in the spikes to maintain the car in its allotted lane via magnetized metal spikes implanted in the road.

The Tsukuba Mechanical Engineering Lab improved GM's self-driving automobile in 1977 by utilizing cameras connected to a computer to drive it across the road at 20 mph using image data processing. Daimler and Mercedes-Benz, two of Germany's major vehicle manufacturers, worked on the VaMoRs idea ten years later.

The VaMoRs was a Mercedes-Benz van with cameras and other sensors that interacted with an onboard computer.

The computer on board is capable of driving the car on its own. This technology was a significant advancement over Japan's autonomous vehicle, since it could go at 56 mph on any road or highway without colliding with other vehicles or obstructions.

This notion was sparked by a $1 million prize offered by DARPA in 2002 to any researcher who could construct an autonomous car capable of travelling more than 142 miles.

Unfortunately, the goal was not accomplished, and the autonomous vehicle that travelled the farthest distance only made it 8 miles. In 2009, Google

launched Waymo, a project aimed at developing self-driving automobiles.

Figure 3: When auto pilot mode is activated

By 2017, Google had produced a self-driving vehicle prototype that had travelled over 2 million kilometers. Consumers are being trained to rely on automobile technology, such as sensors to avoid collisions and assist with parking, in order to trust self-driving cars. In January 2019, Musk claimed that Tesla would phase down autopilot and unveil its first-ever autonomous technology in 3–6 months.

Figure 4: Proto type of autonomous car technology

Tesla owners will be able to drive from Los Angeles to New York City without ever touching the steering wheel before the end of the year. That's a distance of nearly 3,000 kilometers! It will be fascinating to see if Musk follows through on his promise.

Tesla Motors isn't the only corporation making such promises to the globe; in fact, more than half of all automakers worldwide plan to build autonomous vehicles starting in 2020 and continuing into the future. As a result, in the coming years, we may witness a significant growth in transportation computerization.

IV. HIGHLIGHTS OF SELF DRIVING CARS

The use of assistive computer technology in automobiles, such as GPS, cameras, stability control systems, and assisted brakes, has been shown to improve passenger safety and driving quality.

The Navia has previously been used in Northern Australia. Self-driving car. The Navia has a peak speed of 12 miles per

It costs $250,000 and may be purchased commercially, albeit it can only be used in certain settings, such as resorts. The Navia is safe for its intended use in resorts because of its modest speed, which allows it to stop when unforeseen obstructions are encountered.

Figure 5: Shows how to change driving modes

Almost every country has a government agency responsible for lane security, driver licensing, and car registration. The Department of Motor Cars (DMV) is responsible for granting driver's licenses, registering vehicles, issuing registration plate numbers for traffic checks and police surveillance, as well as traffic and general road security for both motorists and pedestrians in the United States.

Figure 6: Shows how Autonomous cars detect surroundings using sensors

Even with the strictest traffic rules and regulations in place, about 6 million car accidents occur each year in the United States, injuring approximately 3 million people and killing more than 90 people on average.

This is where autonomous vehicles come in useful, as scientists anticipate that in the near future, they will dramatically reduce, if not eliminate, automotive accidents. Self-driving vehicles, according to engineers, will have a significant influence on our economy and throughput.

The Autonomous car would also allow entrepreneurs to produce things for an altogether new market, perhaps lowering unemployment and raising living standards. Because Autonomous vehicles can communicate with one another, they can determine the optimum route to take and lessen traffic jamming. This will be beneficial to the environment since fewer pollutants will be created.

The automobile would be more accessible to a broader audience, including non-drivers, those with disabilities, and others who are unable to drive for various reasons. Caregivers and even moms who rely on their vehicles for transportation may feel less stressed and anxious as a result. This might also save money by lowering the amount spent on public transit and taxis.

Autonomous vehicles have been in the works for quite some time. They've arrived and are on their way, which is great news. They will, without a doubt, increase the safety and reliability of our transportation networks. It'll only be a matter of time before we can all text while driving and leave the driving to the automobile.

V. CHALLENGES OF SELF DRIVING CARS

In 2011, Anthony Lewandowsky was found guilty of causing an accident while driving a self-driving car for Google. Following this, a number of Google executives raised worry that safety procedures were being disregarded while developing self-driving cars under Levandowski's oversight.

After officials indicated that more than a dozen incidents occurred during the development of Google's self-driving cars, the ethical use of self-driving cars was called into doubt. This isn't the first time a self-driving car has collided with another vehicle.

Figure 7: Accident of a self-driving car

A Tesla Model S was put into self-driving autopilot mode in Florida, and the passenger was killed when the car collided with a tractor trailer because it failed to stop in time when the trailer came in front of it. Because all self-driving cars are regulated and developed by the same system, a single weakness in the administration of the system might place the lives of many people in the hands of a hacker.

Because many major cities rely on parking to make money, driverless automobiles that eliminate the requirement for parking spots might deprive many

communities and businesses of a source of cash. If it becomes easy to break into the cars' code, the safety of passengers in these vehicles may be jeopardized, a risk that might put countless individuals in danger. When it comes to cyber security, many large corporations are always at conflict with hackers and/or terrorists.

Losses can be sustained in a variety of ways. People will no longer need taxi or Uber drivers since they will be transported to their location immediately after inputting their address.

As more technology and procedures have been automated, human connection has shrunk significantly, and the introduction of self-driving cars will only further limit our engagement with people. Loneliness, despair, and concern may occur as a result of this.

"Three out of four U.S. drivers would be 'Afraid' to ride in self-driving vehicles," according to a March 2016 Autom3obile Association poll. This shows how difficult and time-consuming the transition from human to robotic, pre-

AI is programmed drivers will be. As a result, technological advancements will not benefit everyone evenly.

However, the harmful consequences of this are not limited to that. When self-driving vehicles become a common mode of transportation for many people, the price at which they are made available will widen the economic gap between the rich and the poor, perpetuating economic inequality.

VI. FUTURE SCOPE OF SELF DRIVING CARS

As previously said, there are a number of potential issues and dilemmas that might occur as a result of the development of self-driving cars, but there are also solutions that can be used to lessen the severity of the situation. One fear with self-driving cars is that they will eventually supplant public transit.

If the government concentrates less on public transportation and instead pushes the use of autonomous vehicles, many routes will be blocked off, maximizing mobility for the rich while those with lower salaries struggle to get about. One possibility is to encourage the use of self-driving automobiles while keeping ethical and societal goals in mind.This will guarantee that the needs of the community are met. Another issue that some people may be concerned about is entrusting their lives to self-driving automobiles. Allowing frequent testing of autonomous cars to ensure that they are safe and efficient might be one answer.

The utilization+ of parking spots for self-driving cars would occupy a lot of land area, which is an environmental problem. During the rollout of self-driving automobiles, a dedicated lane for vehicles that can communicate themselves may be required.

Even though it may present problems with the existing scarcity of space, with the rise of technology, a much more effective way to use space may become obvious.

VII. CONCLUSION

Self-driving Car is the major upgradation in vehicle industry in future. Autonomous cars have been in development for quite some time. The good thing is that they've arrived and are on their way. Sure, they make our travelling safer and more efficient. Nature moves quicker than people. Automation can be of much greater assistance.

VIII. REFERENCES

[1]. Sudhir Bhandari, Bishal Poudel, Self driving car simulation, November 2021

[2]. Dr.T. Manikandan, J.S. Jayashwanth, S. Harish and K.N. Harshith Sivatej, self-driving car, April 2020

[3]. M.Vaishnavi, K.Varshitha, G.Usha, C.mounika, July 2021

[4]. Tobias Holstein, GordanaDodig-Crnkovic, Patrizio Pelliccione (2018), Ethical and Social Aspects of self-driving cars, Retrieved from: https://arxiv.org/pdf/1802.04[3.pdf

[5]. Todd Litman, (2018), Autonomous Vehicle Implementation Predictions Implications for Transport Planning, p.12, p.27, Retrieved from: https://www.vtpi.org/avip.pdf

[6]. Luke Dormehi, Stephen Edelstein (2018), Sit back, relax, and enjoy a ride though the history of self-driving cars, Digital trends (https://www.digitaltrends.com/); at https://www.digitaltrends.com/cars/history-of-self-driving-cars-milestones/, Retrieved on: 16/10/18

[7]. Adeel Lari, Frank Douma, IfyOnyiah (2015), Self-Driving Vehicles and Policy Implications: Current Status of Autonomous Vehicle Development and Minnesota Policy Implications, vol.16, issue 2, p.744, Scholarship law (https://scholarship.law.umn.edu/); at https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1015&context=mjlst, Retrieved on: 18/10/18

[8]. Sean Wolfe (2018), A Google self-driving car reportedly caused a crash in 2011 after a former engineer changed its code to drive where it wasn't supposed to, uk Business Insider (http://uk.businessinsider.com/); at http://uk.businessinsider.com/anthony-levandowski-google-self-driving-car-crash-2018-10?r=US&IR=T, Retrieved on: 16/10/18

[9]. Marc Weber (2014), Where to? A History of Autonomous Vehicles, Computer History museum (http://www.computerhistory.org/); at http://www.computerhistory.org/atchm/where-to-a-history-of-autonomous-vehicles/, Retrieved on: 18/10/18

[10]. Futurism (2018), 7 Benefits of driverless cars, Retrieved from: https://futurism.com/images/7-benefits-of-driverless-cars/

[11]. Phil LeBeau (2018), GM tops Tesla in ranking of automated driving systems, CNBC (https://www.cnbc.com/world/); at https://www.cnbc.com/2018/10/03/gm-tops-tesla-in-ranking-of-automated-driving-systems.html, Retrieved on: 19/10/19

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Machine Learning for Phone Call Phishing Detection

Manjunatha B.N[1], Rekha M.S[1], Kapu Pavan Kumar Reddy[2], Kadiyala Charan Kumar Reddy[2]

[1]Assistant Professor, Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Doddballapur, Karnataka, India

[2]UG Student, Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Doddaballapur, Karnataka, India,

ABSTRACT

Phishing is the process of steeling the personal information of the victim. Email phishing is one the major phishing techniques are used by intruders. the intruder's sends the malicious email to the users to trick them by clicking on the link & to make them to enter the personal details like Aadhaar number, bank account details, pan card details, user name & passwords, private personal information etc. However, email phishing is not only the technique used by the intruder's, they also use text messages, fake phone calls, physical connections with victim. There's also a big difference between how people think they'll react to phishing efforts and how they really react. There are some significant techniques are developed to detect the phishing one of them is detection using machine learning.

Keywords— Phishing, intruders, detection, phishing attacks.

I. INTRODUCTION

The use of internet in our daily lives is continuously rising. Technology is likewise advancing at a break neck pace. It became a vital part of our daily lives when compared to recent days, phishing attacks were less even before the outbreak people are duped by the intruder's (or) phishing everything went online as a result of the outbreak and social networking sites have grown significantly in popularity in recent for the attackers to gain access of any system[1]. They focus on people rather than focusing on system to get an access by duping the people. The phishing may be by the text message, phone call, email. The online payment is the most effected when compared to the other industries (or) domains[1][3].

When the email (or) text message received to the devices, people are excited to read and responds to that mail (or) text message by entering the confidential details (or) personal details and when a user ignores a malicious email, the attack is killed instantaneously and no data is lost[1][2][7]. Phishing frauds might be most wide spread cyberattack used today with in the part of the phishing attack the intruder's gets their personal data. Like account numbers, usernames, passwords, etc. Due to the vast nature of phishing attacks, the attack detection identification begins by defining the matter and analyzing the phishing life cycle and effects. detection

of this phishing website is predated to black listed URL. Many other approaches for detecting phishing attack exist like black list, white list, based approach [1], machine learning approach cantina-based approach etc. Phishing websites appear to be valid and many users have difficulty in identifying the fraudulent website [6].

II. LITREATURE SURVEY

In the present market the there are many techniques are available for phishing. Some of them are e-mail phishing, spear phishing and whaling. A fake e-mail will be created to send to the victim (Wosah Peace Nmachi and Thomas Win, 2021). The intruders also create the fake websites to attract more victims to visit and to enter more personal information in the website, so that the intruders can misuse the collected details(P.Kalaharsha and B. M. Mehtre, 2021). The recent website are using the HTTPS to secure the websites, but the intruders are HTTPSifying the phishing websites. Furthermore, because they would need to get a public-key certificate issued for the construction of the websites, adopting HTTPS might contribute to the generation of intrinsic footprints and give defenders with a fantastic chance to monitor and detect websites, including phishing sites.(Yuji Sakurai, Takuya Watanabe, Tetsuya Okuda, Mitsuaki Akiyama and Tatsuya Mori, 2021). To assess these assaults by determining the current status of phishing and examining current phishing strategies. The relevance of the end-to-end lifespan of phishing has been overlooked in studies that have classed phishing assaults according to core phishing mechanisms and counter measures.(Zainab Alkhalil, Chaminda Hewage , Liqaa Nawaf and Imtiaz Khan, 2021).

III. TYPES OF PHISHING

When it comes to the phishing there are many different types of phishing attacks based on their implementation[1]. The most familiar phishing types are like, sniffing, Email phishing, pharming, pop up phishing, water hole phishing, HTTP's phishing Evil Twin phishing, Host file phishing, SMS phishing, Link manipulation, data shop lifting, the below figure exhibit the types of phishing attacks[1][2].

Figure 1: Types of phishing.

In order to clone the webpages of popular websites, the intruder/phisher must put in a lot of effort. To clone the pages, the phishers use web page cloning. The URL and acceptable location will be relative, and the victim will be sent to bogus websites where they will be requested for personal information. The user will then submit the username, password, and/or personal information. The letter includes a link that appears to be from a legal website (or) application. The phisher will make use of the information entered by the user. Phishers have recently taken advantage of the pandemic to deceive the public.

Malware phishing is a type of phishing assault in which the victim or user receives a message (or) email containing malicious code[2]. The code is linked to all of the user's accounts, and when the user clicks the link, the code gathers all of the data from the associated accounts. This primarily affects small and

medium-sized businesses that do not update their application software[2].

Keylogger and screen logger assaults are two different types of phishing attacks. The keylogger will always record the user's keystrokes. It will record without the user's awareness, and subsequently the phisher will steal the victim's passwords and usernames[3].

In email/spam phishing, the message or email is sent to the victim, who believes the message and clicks on the link provided in the mail/message, where the user is asked to enter personal information or update information in order to obtain the subscription credentials. To purchase a membership, the user must submit their credit card and personal information, which the phisher then obtains. In web-based delivery assaults, the phisher detects all of the facts that are transferred during a transaction between the real website and the user, and afterwards transfers all of the information obtained by the phisher without the user being acknowledged. And the phisher will misuse it. The intruder replaces some or all of the original website's content in a content-injection phishing attempt. With malicious softwareto deceive the user into freely providing classified information to the intruder, all data captured will be sent to the intruder's computer (or) to the server[2][8].

The Man-in-the-Middle is a fictional character who lives in the United States The intruder places himself in the middle between the user and the server in a middle-of-the-road assault. The intruder will not alter the data, but he will record all communications between the user and the server. By decrypting the packets being sent from the server to the user, the intruder can acquire the user's personal information[7].

Phishing built for search engines is a sort of phishing in which phishers make websites more appealing to grab visitors' attention. The phishers will add additional offers to the websites so that users would use them and share them with their friends, allowing the phishers to trick users into purchasing products by inputting credit/debit card information, allowing the phisher to abuse the information.

Session hijacking, also known as cookie hijacking, occurs when a phisher takes advantage of the user's session. Session hijacking attempts to obtain access to a legitimate session key in order to steal (or) execute faux authentication on a remote server in the place of a valid user[1].

Web trojans are malicious programmers that infect a user's computer through legitimate channels such as emails, respectable URLs, and so on. The attacker uses this method to place a time bomb that may accidentally detonate when the user attempts to view a file, browse the internet, or perform other ordinary actions. There, trojans capture and communicate the user's local records to the phisher. In a host file poisoning attack, when a user types in a URL to access a website, the phisher must intercept the URL before regular communication can begin[1][2].

IV. STATE OF THE ART

The below figure explains the how the victim gets defrauded by the phisher , when the phisher sends the mail or message to the victim. Either trusting or not its completely depends on the victim. If the user didn't believe the text then the user is safe.

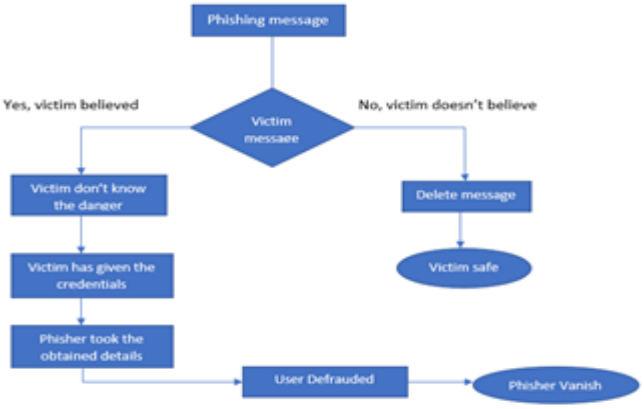


Figure 2: Process of getting defrauded by phisher.

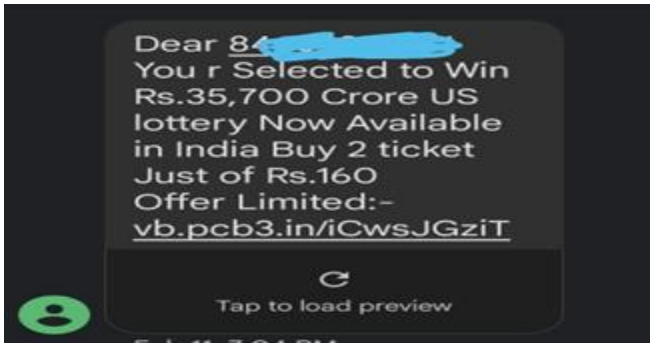


Figure 3: Example for phishing message

According to the fig.(2), When the victim receives the mail/message regarding the subscription expired or any other offer, if the victim doesn't believe that then the victim is safe and the message will be deleted. What if the victim believed that it is true then the victim starts visiting the website and user don't know the danger behind that message. As soon as the victim clicks on that link and enters the personal information like usernames, password, credit/debit card details, address, pan card details, Aadhar number etc.

Immediately the details will be sent to the phisher so that the phisher collects the details and user will be defrauded. In this process not only, the personal details but also many people lost their images, videos and money. Phishers never reveal themselves after losing the details users can't track the phisher. So, the prevention is better than cure. So rather than trusting the messages from the unknown people, just ignore the messages. Even everyday we will get many mails regarding the offers and loan approved to claim please click the link below. The phishers are attracting us towards the messages and mails to click the link which is given in their messages or mails.

The threat has now evolved into a whaling attack, in which an attacker poses as a senior member of the organization and manipulates other members into disclosing sensitive or financial information about the company. The goal is always to steal money or personal data. The suggested scheme examines existing attack data to identify phishing scam behavior and attempts to avoid future attacks by banning such unsolicited emails/messages

V. PHISHING TEDECTION TECHNIQUES:

There are different types of detection techniques for phishing sites. The below figure(4) shows various types of phishing detection techniques.

LIST BASED:

This method has two types:

1.Black listing

2.White-listing

These are also called as traditional approach or the database-oriented approach.

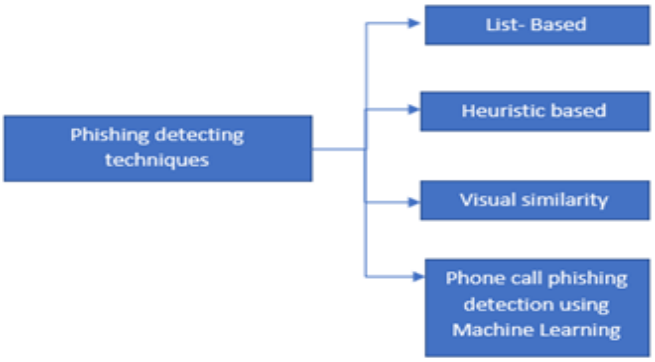


Figure 4: Different Phishing detecting techniques.

In the black listing technique, the URLs that are regarded to be phished sites are recorded in a database, and when a new URL is input, it compares it to the URLs in the database, and if they match, the browser blocks it and the URL is saved in the database for future use. This approach has the drawback of being unable to identify zero-hour phishing assaults. In the White listing technique, the legitimate URLs are maintained in a database and used to validate new URLs using this approach. When a new URL is input, they first search the database for a record of that URL; if there isn't one, the whole information of that URL is examined, including domain names, age, SSL certificates, and hyper-links related to the website, and they are then recorded in the database. The

disadvantage of the whitelisting approach is that it does not ensure that the websites listed as genuine are indeed authentic or that they present themselves as legitimate. The drawback of listing strategies is that they take up a lot of space.

HEURISTIC BASED:

It is similar to the list-based technique. In this technique we extract the URL's, content and they are used in the different sites, if the URL's matches then it is considered as the phishing site, if not then it is good.

This listing technique is far better than the list-based it will give the accurate results but it is too slow. The URL-based detection algorithms are becoming more prevalent as a means of enhancing detection speed.

If URL-based characteristics paired with machine learning improve accuracy, In content based websites the content of the websites is compared to the content of authentic websites to determine if they are real or not. However, there are numerous websites with little substance, therefore this identification approach fails. Images are now used to substitute webpage content.

VISUAL SIMILARITY:

This method involves taking screenshots of online pages and storing them in databases. If there are similar websites, the screenshots of both webpages are compared to see if it is a phishing website or not.

The disadvantages of this strategy are that it takes longer to execute, making it impractical. It necessitates a huge amount of storage space to save webpage screenshots.

When two or more websites with the same URL exist, the first one is assumed to be authentic. However, there is a potential that the first site to show is a phishing site.

PHONE CALL PHISHING DETECTION USING ML:

There are different types data-sets are retrieved from the different features of the websites, machine learning will be trained on the datasets on which they retrieved from the websites and tests them with different machine learning classifiers such as Random forest classifier, Support Vector machine, Decision tree, Naïve Bayes, logistic regression and so on[2].

In recent years, phishing has taken on a variety of forms. Because of the novelty of the attack methodologies, such attacks defraud not only inexperienced users, but even well-educated people. The most common way that cyber criminals target their victims is through business email compromise (BEC). There are numerous approaches for identifying assaults after they have occurred. In this paper, we've talked about a few of them. The general public must be educated to avoid opening and providing critical information to all[1].

The below figure explains how the phone call phishing can be detected by using ML.

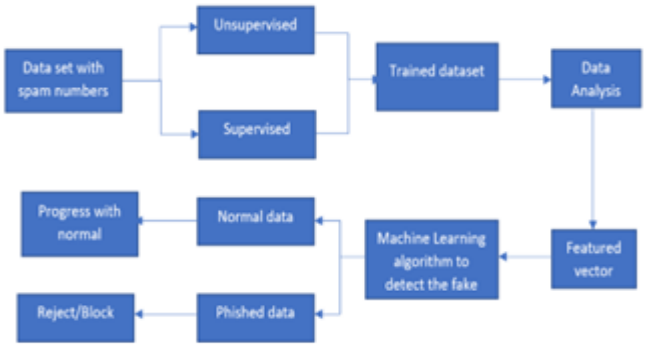


Figure 5: Proposed phone call phishing attack detection

From the figure 5, we have seen how the machine is trained by using the data sets. Here we have the dataset with phishing phone numbers in which we will use 70% of the data to train the machine and the remaining datasets are used to test the system whether the system is trained properly. In which the dataset will have both supervised and unsupervised data. Where we can get the trained dataset for the data analysis. Then we will write an algorithm to detect the phishing phone numbers.

For several years, telecommunication fraud has been inflicting substantial financial damage to

telecommunication customers in India. The most common method for detecting telecommunication fraud is to create a blacklist of fraudulent phone numbers.

Attackers, on the other hand, can easily avoid detection by changing their phone numbers, which is quite simple to do with VoIP (Voice over IP). To address this issue, instead of relying solely on the caller's phone number, we detect telecommunication scams based on the substance of the conversation. We collect telecommunication fraud descriptions from press stories and social media in particular.

Now a days many applications are developed to detect the spam calls. All the users are need to aware about the fraud calls from the phisher. The users need to remember that no bank or no credit card services will ask for OTP. The major reason for the financial loss is without knowing the other person, the victim will reveal the OTP received to the phisher, so that the phisher will theft the amount from our accounts. To escape from the attacks like this we need to avoid the calls or we need evade revealing the OTP, even the bank manager should not ask for the customers OTP. By taking the required precautions the financial loss can be reduced.

Table 1: Shows the accuracy and metrics of different techniques.

S.NO.	APPROACH	TECHNIQUE	METRICS	ACCURACY
1.	Heuristic based approach	Decision tree algorithm	False positive: 5 True positive:120 False Negative:3 True Negative:72	96.76%
2.	Blacklist approach	Simhash algorithm	False Positive:0	84.36%
3.	Cantina based approach	TF-IDF information retrieval algorithm	False Positive: 6%	97%
4.	Machine learning approach	Machine learning algorithm	False Positive: 1.52% True positive:98.39%	>98.4%

The above figure(6) shows about the accuracy and the metrics of the different phishing detection techniques . that will explains what are the true positives and negatives for the different techniques. Heuristic based approach have the accuracy of 96.76% followed by black list approach where it has an accuracy of 84.36%. The machine learning algorithm has the highest accuracy of >98.4%.

VI. CONCLUSION

The different forms of phishing not only impact the normal people but also a well-qualified educated people are scammed by such attacks. Whaling and Spear phishing impacts medium to big organization's by mining all the private and very sensitive data of the organization. Different ways of implementing the phishing attacks have been discussed.

The different ways of identifying the true phishing attacks have been discussed. It became impudent that anti-phishing techniques be in-built with frequent updating to enhance security of every user. After implementing the machine learning on the phone call fraud detection, it will detect the fraudsters to get caught and it will detect some key like asking the phone OTP, account numbers, personal details etc. In the future may have the huge scope in detecting the fraud phone calls.

VII. REFERENCES

[1]. G. JaspherWillsie Kathrine, Paradise Mercy Praise, A. Amrutha Rose, EligiousKalaivani. C, "Variants of phishing attacks and their detection techniques" , Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.

[2]. P. Kalaharsha ,B. M. Mehtre, " Detecting Phishing Sites - An Overview", Center of excellence in cyber security, Institute for

Development and Research in Banking Technology (IDRBT), Hyderabad, India.

[3]. Wosah Peace Nmachi, Thomas Win, "Phishing detection techniques : a literature survey" , International journal of Network security and it's applications (IJNSA) VOL-13, No-2, March - 2021.

[4]. Yuji Sakurai, Takuya Watanabe, Tetsuya Okuda, Mitsuaki Akiyama, Tatsuya Mori, "Identifying the Phishing Websites Using the Patterns of TLS Certificates", Journal of Cyber Security and Mobility, Vol. 10 2, 451–486. doi: 10.13052/jcsm2245-1439.1026

[5]. Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, Imtiaz Khan, "Phishing attacks: A Recent comprehensive study and new anotomy " ,published: 09 March 2021 doi: 10.3389/fcomp.2021.563060

[6]. Kuldeep Singh, Palvi Aggarwal, Prashanth Rajivan, Cleotilde Gonzalez, "What makes phishing emails hard for humans to detect? ", Proceedings of the 2020 HFES 64th International Annual Meeting.

[7]. FarkhondehHassandoust, Harminder Singh, Jocelyn Williams, "The Role of Contextualization in Users' Vulnerability to Phishing Attempts", Australasian Journal of Information Systems Hassandoust, Singh & Williams 2020, Vol 24, Research Article.

[8]. Ms. Neha R. Israni, Mr. Anil N. Jaiswal, "A Survey on various Phishing and Anti Phishing Measures", International Journal of engineering research and technology, Volume.4(Issue 01), ISSN: 2278-0181, January 2015.

[9]. Dipesh Vaya, Sarika Khandelwal, Teena Habpawat, "A Review on Visual Cryptography", International Journal of Computer Applications, Volume.174 (Issue 05), ISSN: 0975-8887, September 2017.

[10]. The biggest phishing attacks of 2018 and what companies can dot prevent them in 2019, available at: https://www.techrepublic.com/article/the-biggest-phishingattacks-of-2018-and-what-companies-can-do-to-prevent-themin-2019/

[11]. What is whaling attack, Available at: https://meen.kaspersky.com/resource-center/definitions/what-is-awhaling-attack

[12]. Himani thakur, Dr.Supreet Kaur, "A survey paper on phishing detection", International Journal of Advanced Research in Computer Science, Volume.7 (Issue 04), ISSN:0976-5679, July-August 2016.

[13]. SaharAbdelnabi, Katharina Krombholz, and Mario Fritz. 2020. VisualPhishNet: Zero-Day Phishing Website Detection by Visual Similarity. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20). Association for Computing Machinery, New York, NY, USA, 1681–1698. DOI:https://doi.org/10.1145/3372297.3417233

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Gesture Controlled Robotic Car Using Bluetooth Module

Sreelakshmi Tadigotla, Samarth Hegde, Nishanth R, Rohit V, Shashank M L
Department of Electronics & Telecommunication, Sir MVIT, Bengaluru, Karnataka, India

ABSTRACT

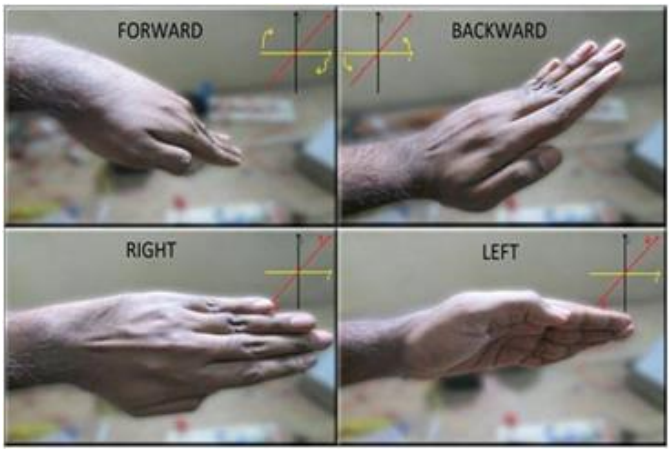
This project is a model for monitoring in real-time that makes the advantages of gestures to allow humans to engage with robots. This is a huge help for folks who have difficulty moving around. Speech recognition has failed to dictate robots because of modulation and variable frequency; hence a vision-based interface is required. The implementation is accomplished through the use of various gestures to guide the robot. Physically challenged people's lives are made easier as a result of this project's impact. It will assist a variety of domains as a result of additional research, including military and high-security bases. Interaction between people and robots is an adaptable human vehicle interface. Working robots will collaborate with humans, making work easier and less complicated. Robotics plays a vital role in residential applications as well as in commercial applications to complete various complex tasks. Many studies are being conducted to improve the human-robot intervention. The main goal of this project is to create a robotic car using Arduino and control it using gesture-based information. The project's entire mechanism is built around the Arduino chip. The "hand gesture" feature allows it to move wirelessly based on the Radio Frequency that is placed on the user's hand.
Keywords-Robotic Cars, Gesture Control, User Interface, Arduino Accelerometer, Gyroscope.

I. INTRODUCTION

Internal circuitry directs a robot, which is a type of electromechanical equipment. Robotics is a new burgeoning industry that will be extremely beneficial to humanity in the next years. Hence, the need for robots is growing across all areas, including industry, education, hospitals, transportation, restaurants, and many others. The traditional method of controlling robots is obsolete. Because of the bulky design of the robot and the long wires used to control it, it is less efficient. In this research, the movement of the hand is used to operate a robotic vehicle. As a result, the robotic vehicle can only move by only hand movements.

Furthermore, the proposed robotic vehicle's control system is developed wirelessly, making it more efficient. In a Gesture Controlled Robot, the controlling action depends on various hand gestures instead of traditional buttons. It is required to carry only a small transmitting gadget with an acceleration meter in your hand. This will send the robot the right command, allowing it to do whatever we want. The transmitting device comprises a converter for analog to digital conversion and HT12E, an encoder IC for four-bit data encoding, which is subsequently transmitted via an RF Transmitter module. At the receiving end, through the RF receiver module the encoded data is received. And then it is decoded by the decoder. The data is then analyzed by a

microcontroller, and the motors are controlled by our motor driver. To make a complex task to simple, divide it into various modules so that the task becomes simple, easy, and error-free. As a result, a person can only move their hand to control the movement of this robotic vehicle in forward, backward, left, and right directions.



Furthermore, live monitoring or surveillance has become a significant concern in our daily lives, with demand growing rapidly. To meet that demand, our proposed method may be utilized to dispatch the robotic vehicle in any direction while capturing live video footage surrounding it using a camera attached to it. Interactions will be easier if humans can completely control machines with their gestures. This suggested system will have a wireless live streaming feature, allowing the user to view anything from their own computer. There are some dangerous locales or situations where this method, rather of sending a human, will be more effective. Furthermore, there are several advantages to using a robotic system rather than a human. For example, a robot can work nonstop in high-pressure, high-resistance, or even dangerous environments.

II. BACKGROUND AND LITERATURE REVIEW

The authors VIJAYALAKSHMI S & ARCHANA M [1] have proposed a temperature-sensing robocar controlled by an android. In their work, they have used various components such as Arduino, a dual-channel H-bridge DC motor along with the driver, Bluetooth modules, and temperature sensors. Arduino is an open-source prototyping platform that may be used to measure the temperature of the environment. By having programming knowledge Any Android device can be used to control this robocar. Air Droid is an Android-only program that allows users to connect the handset to a PC via a wireless network's Wi-Fi module. The Air Droid app is also used to track the car's whereabouts. The main function of this robocar is to perceive the military force's environment before performing certain tasks.

Another work [2] investigated a rudimentary robotic chassis that can be controlled using an accelerometer instead of buttons. The accelerometer is the key device and is a three-axis estimation device with a range of +-3g. This gadget uses a polysilicon surface sensor and a signal controlling circuit to measure acceleration. The output of the accelerometer is analog and corresponds to acceleration. This apparatus monitors the static acceleration of gravity when we tilt it. As a result, there is a moment or vibration. The hand position is detected, and the resulting coordinates are used as the parameter, with the statement being triggered if the required conditions are met. By executing the Arduino code's statement, as a result, the direction of the robot chassis is adjusted. So that it can accomplish tasks like moving ahead, backward, left turn, right turn, and stop. Most of the gadgets are complex and controlled through buttons and switches. Defense, industrial robotics, civil vehicle part assembly, and medical surgery are some of the key sectors that use hand gesture robots. If we use remote controllers and buttons in these fields, the process will become more difficult. Because the operator may become confused between button control and switch, new technology is employed to

control the motion of the hand as well as to manage the device's motion.

Authors in the paper [3] have invented a GCUI prototype program to help users with everyday chores such as making phone calls, controlling their television, and performing mathematical computations. According to a review of studies published in this study over the last 30 years, Gesture Controlled User Interfaces (GCUI) currently offer practical and cheap options that may be suited for the elderly and disabled. Open Gesture is a television interface that employs basic hand gestures to execute a variety of functions. This article assesses the usability of open gestures. In the concluding remarks, the authors have mentioned the ideas to solve the hurdles, this overcomes the dependency and improves the quality of life of elderly and impaired people, as well as for normal people.

A work [4] offers a GCUI model that identifies the latest trends in technology, its usability, and applications in the article Modeling and Designing of Gesture Control Robot. It's an integrated approach for real-time detection and gesture-based data that uses the user's motions to control vehicle movement and manipulation. An accelerometer with three axes has been modified. The accelerometer moves in lockstep with the person's hand movement. The accelerometer records the gesture, then it is processed. Human-machine interactions are becoming more relevant and compatible with the physical world as they move away from the mouse and pen. New technology always improves the level of living; the gap between machines and humans is narrowing.

In this paper [5], the authors have presented that the world is entering a smartphone era in which everything in our daily lives is controlled by a smartphone. Smartphone usage is growing exponentially and is used to control various daily things nearby as well as around us, such as lighting, televisions, sound systems, and air conditioners, among other things. So, in this smartphone-based era, the work intends to build a toy car to entertain children that can also be used to teach them about the world around them is possible to control by only one smartphone app. The authors considered byte-crushing processors with a high-end operation that allows the user to operate the toy car that will play with the child and includes some extra functions that cause entertainment.

In the work [6], The combination of hardware and software developed for completing particular duties for users and controlled by hand gestures should execute the work of the gesture-controlled robot. They have suggested a robot for handicapped and aged individuals in this study. Using motion sensors, the robot identifies five different hand actions. The five hand motions used to govern five directions are reverse, forward, left, right and steady or stop. In this work, they have used motion sensors to identify the gestures which make the system basic as well as efficient. The model can be converted into a wheelchair, a folding bed, or other physical structures with more than two wheels. The user will hold an Arduino Nano along with a gyroscope and an accelerometer. The motors receive the signals wirelessly which will move the structure in reaction to the user's hand motions, allowing them to move around in their surroundings. Hand motions are properly classified using machine learning.

Another work [7] introduced a gesture-based interface for Human-robot interaction. Service robotics is a burgeoning field of robotics study with tremendous societal implications. Service robots directly engage with the people, while previous research has mostly focused on control and navigation, only a few robotic systems are equipped with flexible user interfaces that allow the controlling action of a robot in a "natural" manner. This work

elaborates a sign-based interface for controlling a manipulator-equipped mobile robot. The arm motion movements are observed by using a camera. The robot can accurately follow a person and track him around office surroundings with various intensive conditions due to the adaptive tracking system. The two approaches for gesture recognition are, a neural network method and a template-based method, and a neural network approach are differentiated. Both are used in conjunction with the Viterbi algorithm to recognize motions defined by the arm. The findings are presented as part of an interactive clean-up session in which a person directs the robot to certain cleaning locations and commands it to pick up rubbish.

III. OBJECTIVES

1. Affordability:

The moto behind this work is to design a cost-effective and simple gesture-controlled robot that can be used for various purposes to make life simple.

2. Connectivity and Communication with physical devices:

The use of Bluetooth allows human-machine interaction and it is possible to communicate between the devices.

3. Intelligent Sensing abilities:

Sensors such as accelerometers can sense very minute movements, for instance, little vibrations, which humans cannot even recognize.

4. Efficiency:

It is possible to identify a very little movement on a very large scale. This reduces human energy and provides maximum work.

5. Simplicity:

By eliminating input devices like joysticks, mice, and keyboards and allowing the unfettered body to provide messages to the computer through gestures like finger-pointing, this technology has the potential to transform the way users interact with computers.

IV. METHODOLOGY

This is a simple application for the MPU-6050 3-axis Gyroscope and Accelerometer.

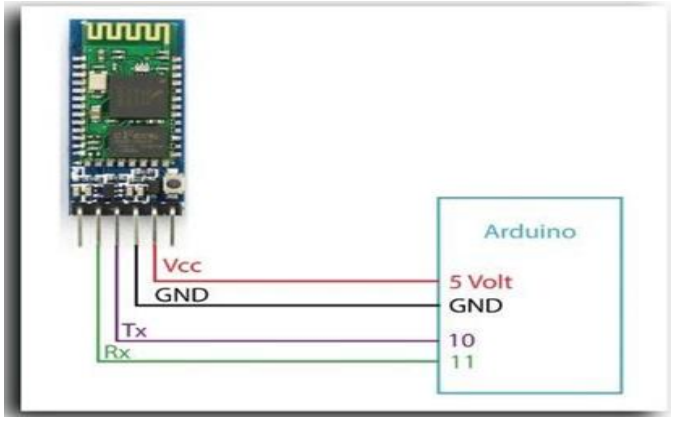
Concentration is primarily on communication between two Bluetooth modules (HC-05).

The factory setting on the HC-05 module is for a slave. That implies simply a connection of Arduino to the module can be done and data can be sent to it.

The data acquired by the gyro sensor MPU 6050 is communicated between two Bluetooth modules.

To begin, we must first configure these two Bluetooth modules.

One Bluetooth module serves as the slave, while the other serves as the master.



The second Bluetooth module is configured as a Master device and connects this master device with the Arduino board and enters it into AT mode as done previously. Using a USB cable connects the car with the PC. By selecting the proper com port and board type, upload the program. While uploading the program, the battery and Bluetooth module should not be connected.The remote unit should be uploaded with the program in a similar manner. Mount the slave module on the car and the master module on the remote unit.

V. BLOCK DIAGRAM

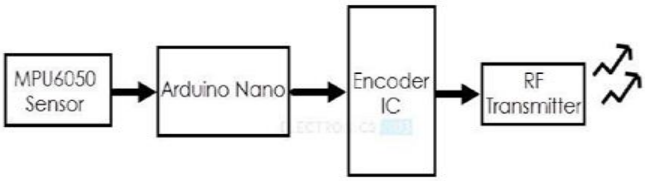


Fig. Transmitter Block Diagram.

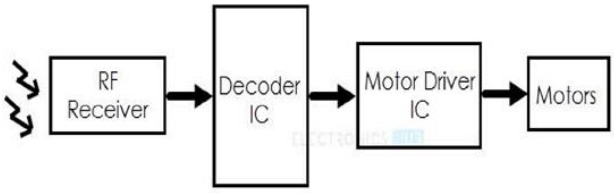
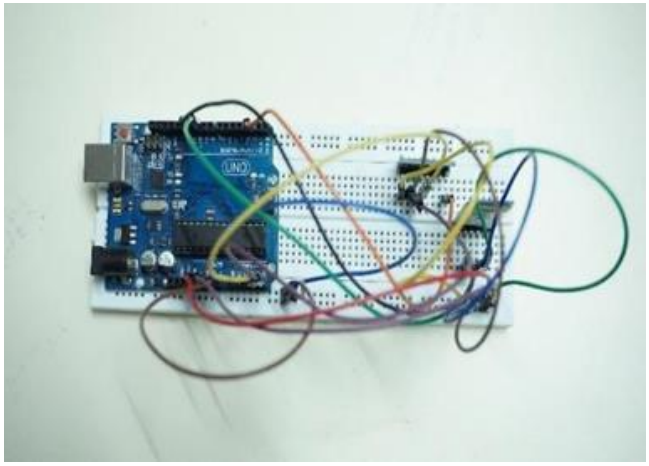


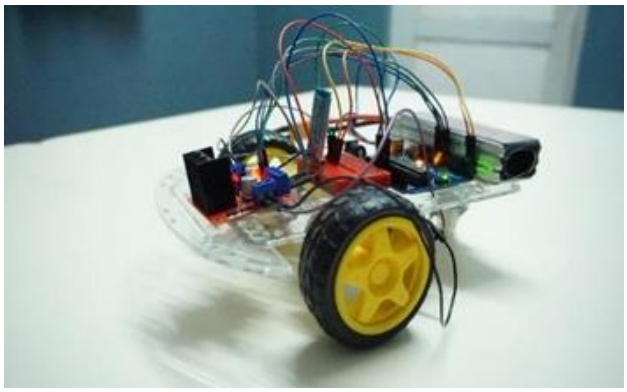
Fig. Receiver Block Diagram

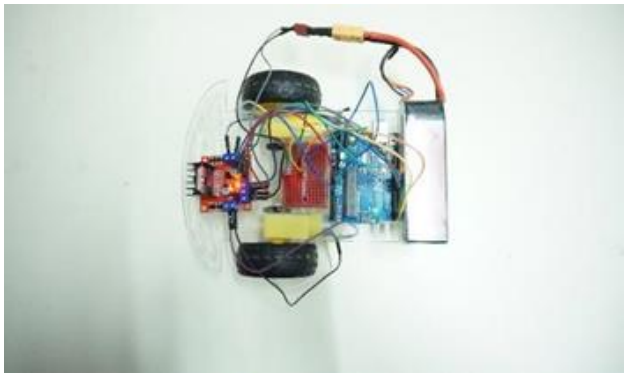
VI. RESULTS AND DISCUSSION

After simulating the circuit connections, the movement of the robotic car is through gestures. The connections were established according to the circuit schematic, and the Arduino code file was uploaded to the Arduino uno.

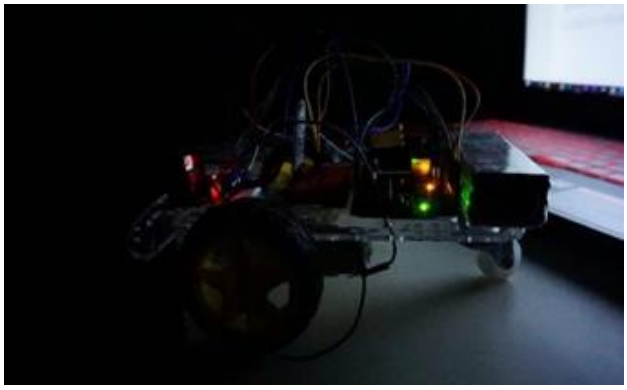


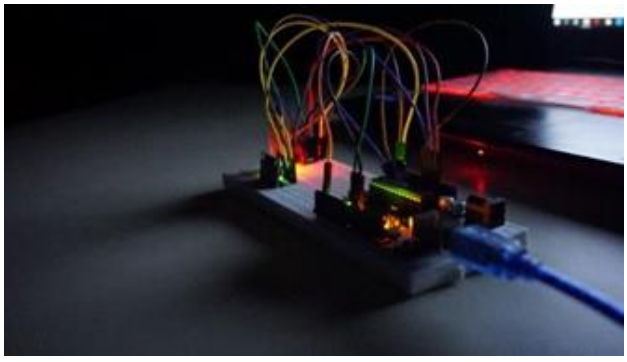
The above picture is of the master unit which contains Bluetooth module-HC05, Arduino Uno and other sensors.

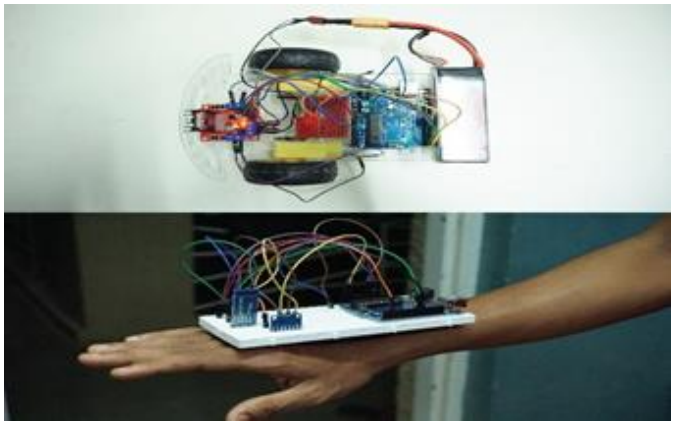




The above Pictures are of the slave unit which contains Bluetooth module-HC05, Arduino Nano, Accelerometer and other components.







The above picture is the final product of the project where master unit is controlled through hand gestures and slave unit responds to the gesture with respective movement.

VII. FUTURE SCOPE

Robotic arms can be designed like human hands in the future. That can easily be controlled using hand gestures only. It can also be used in the field of construction, medical science, hazardous waste disposal, etc.

VIII. REFERENCES

[1]. "Robotic car using Aurdiuno with bluetooth controller", Vijayalakshmi s & Archana m Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India, Assistant Professor, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur Tamil Nadu, India. Journal Publication, Volume 1, Issue 1 | Pages 1-8.

[2]. "Hand Gesture Control Robot" BenjulaAnbu Malar M B, Praveen R, Kavi Priya K P. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2, December 2019.

[3]. "Gesture Controlled User Interface for Inclusive Design and Evaluative Study of Its Usability" Moniruzzaman Bhuiyan and Rich Picking.Journal of Software Engineering and Applications, 2011, 4, 513-521doi:10.4236/jsea.2011.49059 Published Online September 2011 (http://www.SciRP.org/journal/jsea).

[4]. "Modeling and Designing of Gesture Control Robot" ArghaSamanta, Sangharsh Kumar, Deepak Saraswat, Bharnab Das, Vibhuti Kumar, Harpreet Kaur Channi presents a model for gesture-controlled user interface (GCUI), and identifies trends in technology, application and usability. IJSART - Volume 3 Issue 10 – OCTOBER 2017

[5]. "Bluetooth Controlled Car System ", ShrutikaKanere, Rutuja Shinde, VaibhaviTornekar, Dnyaneshwari Zagade, Mrs. V.S.Sonar. International Journal of Advanced Research in Computer and Communication Engineering, ISO 3297:2007 Certified, Vol. 6, Issue 3, March 2017.

[6]. " Designing and Implementation of a Wireless Gesture Controlled Robot for Disabled and Elderly People", MahbubaAlam and mohammadabu Yousuf.Conference Paper· February 2019, 978-1-5386-9111-3/19/$31.00 ©2019 IEEE.

[7]. "A Gesture Based Interface for Human-Robot Interaction" Stefan Waldherr, Roseli Romero, Sebastian Thrun. Autonomous RobotsVolume 9Issue 2September 2000 pp 151–173.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Inkblot Password Security System to Protect Brute Force Attack and Phishing Attack

Apeksha N H, S.J. Subhashini, Manikandakumar M, Asha Rani Borah
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Many security primitives use text-based passwords that are vulnerable to Brute Force and Dictionary attacks. Here a new security primitive is introduced based on a graphical picture authentication system employing inkblot images to solve this challenge. This type of technology protects users from phishing attempts since the user can recall the inkblot image he saw during password setup, and if the inkblot image differs throughout the login process, he can tell it's a phishing site. This security measure was applied on the banking application to improve the system. The new method is simple and relies on a user answering a series of questions when he or she initially registers for access to a website. It starts by randomly placing different inkblot images in a small region of the screen to create a set of random inkblot pictures. The user is required to submit a short word that characterises each of these images as part of the signup procedure. The current system employs text-based passwords, whereas the suggested system employs inkblot images, a sort of graphical password that is more efficient and user-friendly.

Keywords: Password, Brute-force attack, Graphical password, Inkblot images

I. INTRODUCTION

Inkblot simply means a blot of ink Inkblot images are patterns that are created using this blots of ink. The uniqueness of these images is that every human looks at it and interprets it in a different way. This is used as a key factor in Inkblot password security system.

Users are shown the inkblot patterns when they return to the site with a password. Then they must assign the appropriate sentence or a word to each pattern. These are puzzles that are simple for humans to solve but difficult for computers to solve, even if they have the random bits required to create the challenge. This new test is IPS (Inkblot Password Security System).CAPTCHA- Completely Automated Public Turing Test to Tell Computers and Humans Apart is the security method in place at the moment. It's a computer program or system that can tell the difference between human and machine input, usually to prevent spam and automated data extraction from websites. It's a form of challenge answer test used in computing to see if the user is human, but it's also less secure. The disadvantages of present methods include their inability to totally block spam. It is less secure and is vulnerable to Brute

Force, Dictionary, and Optimal Character Recognizer attacks.

II. LITREATURE

[1] Since the HIT's inception, the test's psychometric qualities have been the primary focus. The main goal in this research was to create an alternative to the Rorschach, retaining all of the HIT's positive characteristics while enhancing its psychometric foundation (Holtzman et al., 1961). However, the way HIT variables are coded causes data loss. Previous research has demonstrated the HIT's effectiveness as a research tool and has addressed this restriction by applying the Rorschach coding system to HIT protocols or developing new variables using an HIT coding system. There were some limitations to this investigation. The first is that the information was gathered from a convenience sample of people drawn from the general community. A second drawback is that interrater reliability coefficients for both HIT and Rorschach protocols were not determined for the five separate psychologists who collected and coded the protocols.Despite these flaws, the current study provides a unique viewpoint on HIT research and the test's future. [2] The impact of a person's mentality on the construction of appropriate passwords was investigated in this research. The strength of passwords was one of the issues investigated. Because most password strength metres calculate the entropy of a password rather than its quality, they can only tell users whether a password is weak or strong, not why. The survey that was sent out to diverse consumers also helped in understanding why basic passwords are so popular. The fear of forgetting the password is the primary motivator in the majority of situations. [4,10] The majority of graphical passwords have been found to be vulnerable to shoulder surfing, smear, and brute force attacks. Shoulder surfing attack: this attack includes employing direct observation techniques, such as peering over someone's shoulder, to gather information about the input of other users. It IS especially useful in crowded places because it's reasonably easy to watch someone without being seen. Smudge attack: When screens are touched, oily residues, or smudges, are left behind as a result. Information leakage can be inferred from latent smudges, which can be used to infer recently and often touched portions of the screen. Brute force attack: The goal of a brute force attack is to check all possible secrets until the correct one is discovered. [5]In the area of Internet banking, it is clear that using a single-factor authentication (e.g., a password) is no longer regarded secure. Automated password gathering algorithms are sure to target easy-to-guess passwords like name and date of birth. Two-factor authentication was established to fulfil the needs of enterprises who wanted to give their users with a stronger and safer authentication method. Invalid password attempts should also be kept to a minimum to keep the application safe from hackers, but human error is an issue that might cause problems. Some banks have severe restrictions regarding invalid attempts, which is particularly annoying for e-banking consumers because there is no room for error and is not user-friendly. [11] Most of the online banking applications have one fundamental flaw: if an attacker enters a username that does not exist in the system, the system will either not show any graphic/password or will show a new graphic/password even if the same username is used. Because attackers can attempt as many usernames as they like to see if the username exists in the system before breaking the password, such online banking apps are vulnerable to frequency of occurrence analysis and shoulder-surfing assaults. The attacker can easily shoulder-surf the traditional password input by the users to break the password. As a result,

it is critical to be aware of such a security danger and to suggest a countermeasure to such attacks. [12] In today's world, all internet services demand user authentication. Day by day, governmental institutions and businesses make their services available to the general public over the internet. These governmental entities can determine the best authentication techniques that suit their demands because the findings from matrices are not limited to evaluating online banking. [3] The banking system requires increased security for client data stored in their systems. The security of a client's data is safer with today's technologies. As security improves, the number of frauds rises as well. Many hacking sites have a habit of stealing information from the banking sector. Graphic password strategies are used in various foreign countries. Customers' identities are protected by a graphical password and an OTP. This system should be used all around the world. [6] Graphical passwords are a substitute to alphabetic passwords and can be used to protect online accounts in the same way. The widely used alphanumeric passwords have challenges with memorability, and users often struggle to remember a large number of different passwords. Researchers have used several graphical password schemes (GPSs) to address such security and usability challenges since 1996. There are a variety of such plans accessible. Graphic passwords, on the other hand, are viewed less secure but provide superior memorability. As a result, numerous researchers have looked into graphical passwords as a way to circumvent the flaw. The "Shoulder Surfing Attack," which entails getting into a victim's computer to acquire the entire password, a portion of the password, or any personal information, is one of the most important flaws of graphical passwords. [7] Graphical passwords are more difficult to crack than text-based passwords, and they are easier to remember for users. Research towards graphical passwords that can replace existing text-based passwords has been active in several locations throughout the world in recent years. This article examines recent research and development efforts in Korea related to graphical password authentication systems. For this aim, security authentication techniques based on graphical passwords are divided into technical groups, and research on graphical passwords in Korea is examined. [8] The common phishing attack stories begin with an email requesting an urgent invoice review or password change and conclude with a data breach in which personal information is stolen and money is lost. Although many of us may roll our eyes at the idea of falling for such an obvious con, we must admit that hostile actors would not keep trying if those methods didn't succeed. As an era where we are flooded with online danger, phishing attempts could be regarded a "classic" kind of cybercrime. Although there is no one- size-fits-all approach to phishing prevention and mitigation, security teams may save time and effort by harnessing threat intelligence and implementing tighter filters. [15] A system's security is only as strong as its weakest link. Despite ongoing efforts to educate users about the necessity of password's security and enforcing password creation restrictions on them, the human factor still exists in many information systems. Not only do most people's password design and management habits stay mostly unchanged, but password cracking tools and, more crucially, hardware continue to improve. The passwords used by students at a Slovenian institution to access the online grading system were subjected to a broad targeted attack in this research. The purpose was to show how basic and predictable patterns can be used to hack most user-generated passwords. According to the findings, a single low to mid-range current GPU can crack over 95% of passwords in a matter of days, while a more dedicated system can

crack all but the strongest 0.5 percent of them. [9] For all criteria, the Inkblot and Fractal classes score exceptionally well, whereas the Texture and Snowflake classes perform badly . The bits per character for each image class are nearly identical, showing that response length was the most important element in establishing the image description's security. As a result, there was no difference between individual photos inside image classes in the majority of studies, but there were numerous changes across class boundaries. The next section describes the experiment that tested the durability of descriptions. According to the research, the responses are secure enough to serve as a viable cueing system. The use of inkblots and fractals as password hints is possible. Inkblots are indeed suitable cueing visuals. [13] inkblot is a user-friendly, highly scalable, and secure authentication system that is both simple to use and robust enough to keep malicious users out. Its strength rests in its simplicity and each individual's distinct vision. . [14] The CaRP scheme can be generalized by presenting the CS-AV scheme, which is based on CaRP and uses a suitable symbol selection on the screen by clicking, but does not provide an alphabet. It is seen that combining alphabetic and visual symbols in one alphabet can improve its usability (CS-AV). Techniques can readily identify text-based Captcha, but an automatic brute force attack can only find a graphical password probabilistically . As a result, the CT and CS- AV schemes are compared and find that the CS-AV scheme has a 3.75 percent higher memorability than the CT scheme and is significantly easier to remember than the CT scheme. [16] Replacing passwords with passcodes is another method to improve password security. Pass code authentication appears to be a mathematical calculation, but it is not. As we stated to the hacker, it will be a computation with logical meaning, a time period, and options that will confuse him. For the authorised user, though, it's simply a matter of selecting an item from a list. The user can log into his account without any major analytical progressions or trigonometry by making a simple guess. The main benefit of this authentication is that the user does not need to memorise any passwords. The only thing he needs to remember is the operator's intrinsic meaning. The magic of this verification is that the question shown changes all of the time. As a result, the user may be unsure of the next inquiry. The purpose of this authentication is to eliminate passwords from the user's load. This authentication may be appropriate for small-scale security devices such as mobile phones and computers. By improving, the proposed technique can be utilised to implement in all internet-based logins.

III. METHODOLOGYAND IMPLEMENTATION

A. Existing system

CAPTCHA-Completely Automated Public Turing Test to Tell Computers and Humans Apart is the security method in place at the moment. It's a computer program or system that can tell the difference between human and machine input, usually to prevent spam and automated data extraction from websites. It's a form of challenge answer test used in computing to see if a user is human, but it's also less secure. The current system's drawbacks include its inability to totally avoid spam. It is insecure and vulnerable to Brute Force, Dictionary, and Optimal Character Recognition attacks. Most individuals find audio captchas difficult to comprehend. When a website first starts employing captcha, it notices a significant decline in traffic, and since traffic is money for websites, captcha is costing them money.

B. Proposed System

To address the captcha issues, a mechanism called IPS (Inkblot Password Security) has been developed. It is a graphical password authentication system, which means that instead of numbers and words, we use visuals. IPS, like CAPTCHA, tries to thwart hackers and trolls by introducing a step that is simple for humans but challenging for machines. The photograph will be saved in the database beforehand. Each image is saved with a unique name during the account creation process, and users are requested to provide a description of the images upon registering. When they try to log in the next time, they are requested to provide the same description that they provided upon registration. Only users who meet those descriptions will be able to log in; otherwise, they will be unable to do so. The proposed approach has the following advantages: it is extremely tough to hack the system even if we know the password. Because we use images instead of words or numbers, it provides a high level of security

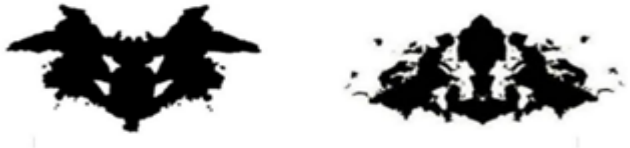


Fig 1: Inkblot images

C. Module Description

Admin:

i. Admin Profile (Add, Edit)
ii. Account Details (Add, Edit, Delete).
iii. Send Email which contains account holder information with inkblot image links.
iv. Branch Details (Add, Edit, Delete).
v. Change password

Admin module description:

Admin can able to view the profile. He has the permission to Add, Edit or Delete any number of users. When he add the new account he has to send inkblot image links to corresponding user email id which contain account number and account holder name. Account types should be like this savings, salary and children's. He will be maintaining the branch details and also he has the permission like add, edit and delete the branch details.

User:

i. Login By using Inkblot images
ii. ii. Account Details (View, Update).
iii. iii. Balance Enquiry.
iv. iv. Amount Transfer Process.
    a. Add Beneficiary Account
    b. Confirm Beneficiary
    c. Transfer Amount
v. Transaction Details.
vi. Change Password.

User module description: During inkblot password setting process user will get different inkblot images and User has to login by using inkblot images.

He can able to view his profile and balance enquiry, if he wants to do any transaction he can to do it

D. System Architecture

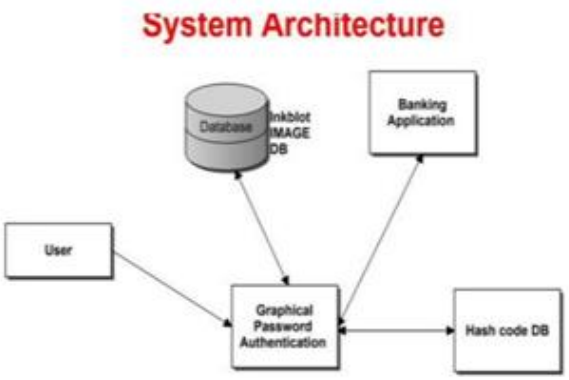


Fig 2: System Architecture

E. Use Case Diagram

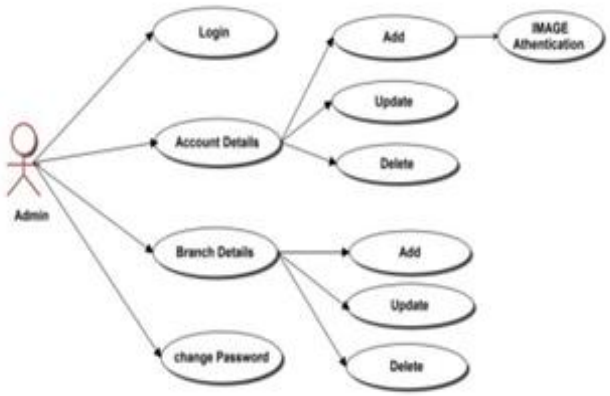


Fig 3: Admin Use Case Diagram

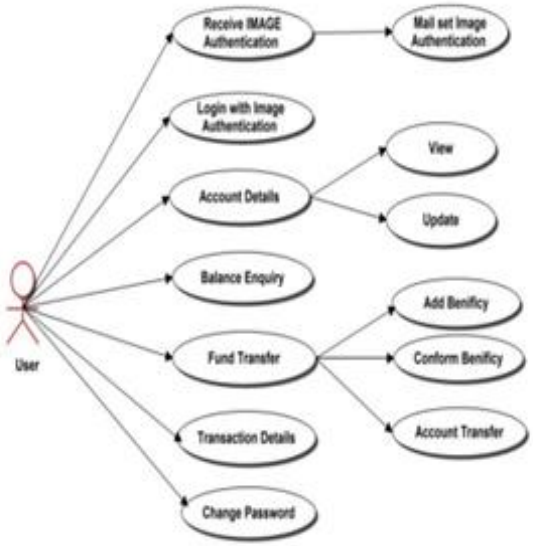


Fig 4: User Use case Diagram

IV. RESULT

The below table shows the comparison between Text bases, Token based and Graphical passwords. It shows how Graphical passwords are better and more user friendly.

Table 1: Comparison of types of passwords

Comparison	Password Authentication Systems		
	Text Based	Biometric/Token Based	Graphical Password
Security	Least	High	Highest
Required Cost	Nothing	Higher	Less
Usability	Easy	Complex	Easiest
Availability	Always	Not Always	Always
GUI	User Friendly / Not attractive	Not user friendly / Attractive	User Friendly / more Attractive

V. CONCLUSION

The study gives a brief picture of how graphical images are better than text passwords and how Inkblot images are even better. Graphical passwords are more human and user friendly. The security of the banking system will be more as the password is more efficient. Brute force and fishing attacks can be avoided. The use of graphical passwords as an alternative to traditional text-based passwords has grown in popularity over the last decade. We did a detailed survey of available graphical password approaches in this work. Although the primary benefit of graphical passwords is that they are easier to remember than text-based passwords, existing user studies are limited, and there is nopersuasive evidence to support this claim. Using typical attack methods such as brute force search, dictionary attack, or malware to crack graphical passwords is more challenging, according to our early findings. Overall, today's graphical password approaches are still in their infancy. However much more research is still required in this field.

I. REFERENCES

[1]. Alessia Michela Di Campi,Ca' Foscari University of Venice Department of Environmental Sciences, Informatics and Statistics, "Password

guessing: learn the nature of passwords by studying the human behavior",2020-2021

[2]. SikiruSubairu Department of Cyber Security Science Federal University of Technology Minna, Nigeria.John Alhassan Department of Cyber Security Science Federal University of Technology Minna, Nigeria. ShafiiAbdulhamid Department of Cyber Security Science Federal University of Technology Minna, Nigeria. Joseph Ojeniyi Department of Cyber Security Science Federal University of Technology Minna, Nigeria, "A Review of Detection Methodologies forQuick Response code Phishing Attacks", 2020

[3]. Mrs. R. Yamini, 2 Mr. S. Ambresh, 3Mr.V. Arjun, 4Mr. M. Vishnu 1Assistant Professor, 2Student, 3Student, 4Student 1 Department of Computer Science andEngineering, 1 Adhiyamaan College of Engineering, Hosur, India, "SECURE INTERNET BANKING USING GRAPHICAL PASSWORD WITH OTP",2020

[4]. Tahir Mehmood, Dr. Ghulam Muhammad Shaikh Department of Computer Science Bahria University Karachi, Pakistan,"Usability and Security in Internet Banking Password (Survey)",2020

[5]. Tahmina Islam Shammee, TaslimaAkter, MuthmainnaMou, Farida

[6]. Chowdhury*, and Md Sadek Ferdous Department of Computer Science and Engineering, Shahajalal University of Science and Technology, Sylhet, Bangladesh, "A Systematic Literature Review of Graphical Password Schemes", 2020

[7]. Akash Rao Old Dominion University, "Cyber Security- A New Secured Password Generation Algorithm with Graphical Authentication and Alphanumeric PasswordsAlongWith Encryption",2019

[8]. Gi-Chul Yang Department of Convergence Software, Mokpo National University Muan-gun, Jeonnam 58554 – Korea,"Development Status and Prospects of Graphical Password Authentication System in Korea",2019

[9]. Adam Vincent, "Don't feed the phish:how to avoid phishing attacks",2019

[10].Karen Renaud & Tony McBryan University of Glasgow, "How Viable are Stubblefield and Simon's Inkblots as Password Cues?" [10](FOURTEEN) FarajiSepideh, "Providing a Secure Hybrid Method for Graphical Password Authentication to Prevent Shoulder Surfing, Smudge and Brute Force Attack",2019

[11].Lip Yee Por , Tan Fong Ang,Department of Computer System & Technology, University of Malaya,Kuala Lumpur 50603, Malaysia and Chin Soon Ku , Department of Computer Science, UniversitiTunku Abdul Rahman, Kampar 31900, Malaysia, "Preventing Shoulder- Surfing Surfing Attacks using Digraph Substitution Rules and Pass- Image Output Feedback", 2019

[12].Anoud Bani-Hania*, Munir Majdalweieha , Aisha AlShamsia , a College of Technological Innovation, Zayed University, Dubai 19282, UAE, "Online Authentication Methods Used in Banks and Attacks Against These Methods",2019

[13].Monika N, S Shree BN Kavitha,Pallavi V Reddy,"IPS: Inkblot Password Security",2018

[14].Altaf Khan(&) and Alexander G. Chefranov Department of Computer Engineering, Eastern Mediterranean University,TRNC, Mersin 10, Turkey, "A New Secure and UsableCaptcha-Based Graphical Password Scheme",2018

[15].L.Bosnjk,J.Sres and B.Brumen,University of Maribor, Faculty of Electrical Engineering and Computer science/Institute of Informatics, Maribor,Slovenia,"Brute- force and dictionary attack on hashed real-world passwords",2018

[16].Prabhakar Gantela, Assistant Professor, Department of Information Technology Mizan-

Tepi University, TepiCampus,Ethiopia. Dr. R. Mahammad Shafi ,AssociateProfessor, Department of Electrical & Computer Engineering, Mizan-Tepi University, Tepi Campus, Ethiopia, "Replacing Passwords with "Passcodes"",2016

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Securing Data Transfer Using Cryptography Techniques

Divya S., Manikandakumar M., S. J. Subhashini, Asha Rani Borah
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

With the increases transmission of data, the data is prone to several attacks in terms of network security. With new advancements in technology, a major challenge that arrives in terms of data is security. The electronic data is transferred through the web and other media. In that case, protection of data can be achieved by cryptography and encryption techniques. By this the normal text or data is encrypted in the form of cipher text. Cipher text is difficult to modify and understand by people. It requires specific technique to decrypt the cipher text and obtain the information. Encryption techniques also use mathematical concepts in data encryption and decryption. In this paper, some of the algorithms like AES, RSA are discussed with reference to file security. The image cryptography by pixel modification technique. And also to make a comparative study for most important algorithms in terms of data security effectiveness, execution time of the algorithm to encrypt and decrypt, key size.

Keywords: Data Security, Cryptography, Encryption, Decryption, Cipher Text, Public Key, Private Key

I. INTRODUCTION

Data sharing has become much easier these days. Data is transferred through various methods like internet, physical connection, and many more. But the security is one of the top priorities of any data transfer. Data can be in any form like a document file, audio file, media file or an image file. In any form the data is to be secured. There are some situations wherein the data must be hidden or secured in other form before transmitting. Especially, in the fields like defence, financial matters, government policies and organisational secrets. In such cases, the cryptography techniques can be deployed for these causes.

The cryptography technique includes various ways like encryption-decryption mechanism which follows some of the mathematical formulae, steganography used to hide secret information in images. Some of the encryption algorithms to be noted are AES, DES, DSA, RSA, ECC, Blowfish, Twofish, and many more. Other ways of securing the files are digital signatures, digital certificates, digital watermarking. Digital signatures and certificates mainly focus on the authorization and security of the information files. Digital Watermarking includes embedding secret information in text, image or numerical form into an existing image. There are two types of encryption-

symmetric and asymmetric. It main difference is the type of key.

II. LITERATURE SURVEY

Omar G. Abood, Shawkat K. Guirguis [1] talk about a brief survey on the cryptography algorithms, and the goals of cryptography are elaborated. Also the Advanced Encryption algorithm(AES) and DES are explained in detail. This gives a brief idea on the algorithms that are designed using the available resources. Kajal, Rizwan Khan [2], introduces the scope / hierarchy of network security technology. The demand for privacy and security has become more these days. Hence, various integration and later extraction techniques are being developed. These include cryptographic techniques and steganography techniques. And the use of cryptography and steganography makes communication more secure. A new method for image steganography which is much more secure than the previous implementations is introduced [3]. Although, it uses some common steganography techniques, integrating it with cryptography and neural networks make it arduous to break. With deep neural networks encryption layer gives adds the security. The image is embedded and revealed the secret image from the container. Further, the main reason to add an additional encryption layer was that if the original cover was made public, the secret will still remain secure. If the additional layer is not added, one can partially decrypt the sensitive information [3]. The DHTTIE is proposed by Ahmad Abusukhon and others. It is a hybrid encryption algorithm which is basically about the Diffie Hellman and TTIE algorithm. The author describes that there are different plans for security in network and cryptography [6]. In cloud, only the sending and beneficiary end would be aware of the secure key. It concentrates about the PC security, key circulation, administration, and cryptographic calculations. The Advanced Encryption Standard (AES) and the Data Encryption Standard (DES) are also thoroughly discussed. They provide a basic summary of cryptography methods, as well as the aims of encryption. The Advanced Encryption Standard (AES) and the Data Encryption Standard (DES) are also thoroughly discussed [6]. The authors talk about asymmetric key encryption. [7] Protection for network-to-network data communication is done by cryptography. It uses data against unauthorised users to protect them. Techniques used are like cryptography, watermarking, digital signatures, firewalls etc. The demand for secure communication has paved path for cryptographic systems become popular. And it has been proved as cryptography is safeguarding our confidential information. The author of the study [8] discusses image watermarking. It is a time-saving method of securely transmitting data. Many watermarking approaches for safe data transmission have been proposed previously. The majority of the work is spent embedding the watermark in the image, but the security of the watermark after it has been embedded in the image is overlooked. As a result, the embedded watermark's security is an important factor to consider when designing a watermarking system. This concept can be developed further to provide security to the watermark image that will be embedded in any media file. The comparison of encryption algorithms is discussed in paper [9.] The parameters of cryptographic algorithms vary, including encipherment and decipherment time, memory, throughput, and CPU utilisation. A survey of the cryptographic algorithms such us ECC, El-Gamal and RSA is conducted. [10] These algorithms are studied in order to be compared. In ECC, the cost of transmission is drastically reduced, as evidenced by comparisons. The outcome demonstrates ECC's useful

performance. Rajani Devi. T., has described briefly about how cryptography works. [11] The reader must beware, however because there are a number of ways to attack every one of these systems; cryptanalysis and attacks on cryptosystems, however, are well beyond the scope of this paper. It represents a survey of the main important cryptographic algorithms. And the choice of keys determines the strength of cryptography. Network security and cryptography algorithms with particular regard to digital signatures is summarized. [12] The transformations SA (SIGNING) and VA (VERIFICATION) are typically characterized more compactly by a key that is, there is a class of signing and verification algorithms publicly known, and each algorithm is identified by a key. Handwritten signatures are interpreted as digital signatures of a special class. The verification function is used to check whether the signature on a particular message done by a user A is actually sA, where sA is the digital signature of A. Applications in the field of information security, authentication, non-repudiation, and data integrity, use digital signatures. One of the prominent application is the certification of public keys in large networks. An overview of cipher texts, procedure of handling cipher texts, types of ciphers and digital signatures. [13] A picture of cryptanalysis, various types of attacks (birthday attack, cube attack), and hash function used for security of the information is obtained. [14] The design and implementation of the digital signature encryption is demonstrated along with the code and interface developed. [15]. The digital secure uses the ECC technique to achieve encryption and security. One of the main applications of encryption is the authentication process, biometric encryption [16]. The protecting aspect of computer network with encryption techniques are also detailed. [17] Image cryptography is one of the techniques these days. It can be achieved by dividing the image into pixels. And the range of pixels which results in least changes is identified and replaced by the text or message is encrypted. Colour image encryption based on an improved Fractional-order logistic map is discussed [18].

The authors brief about neural cryptography based on the CVTPM is investigated. The security and synchronization time of CVTPM are also investigated. It is found that the security of CVTPM is higher than TPM with the same hidden units, input neurons, and synaptic depths. The two parties with CVTPM can exchange two group keys in one neural synchronization process [19]. Bio-Inspired Cryptosystems are a modern form of Cryptography where bio-inspired and machine learning techniques are used for the purpose of securing data. A system has been proposed based on the Central Dogma of Molecular Biology (CDMB) for the Encryption and Decryption Algorithms by simulating the natural processes of Genetic Coding, Transcription and Translation and the reverse processes to allow for encryption and decryption respectively. The proposed bio-inspired cryptosystem shows competent encryption and decryption times even on large data sizes when compared with existing systems. [20]

III. PROPOSED WORK

It mainly concentrates on the module that gives a gives a hybrid encryption and decryption system using DES and RSA. It implements modified version of DES and RSA for file encryption at different levels. The key generated at the encryption end can be shared with anyone, but it cannot be decrypted except for the user whom it was encrypted. And the file must be decrypted in the same mode of encryption. The research aims to perform a study on the comparison of encryption algorithms. To compare the algorithms AES, DES and RSA based on the

execution time, and input size. It gives a comparison of selected encryption algorithms based on the user input for numerical data. RSA is an asymmetric algorithm whereas the AES and DES are symmetric algorithms. The comparison of the encryption algorithms can be done using various factors, out of which some of the major factors areEncryption Time – It is the time taken for obtaining cipher text from plain text. It depends on plain text size. Decryption Time – It is the time taken for obtaining back the plain text from cipher text. Memory used – The memory used for the implementation of the algorithm and the buffers used. Input size – The input size or the length of the data to be encrypted is also considered. It also provides module for image cryptography. It is the art of hiding data within image. By using the pixel modification technique in images.

Algorithms	Blow Fish	AES	3DES	DES
Key size (bits)	32-448	128, 192, 256	112 or 118	64
Block size (bits)	64	128	64	64
Round	16	10, 12, 14	84	16
Structure	Feistel	Substitution Permutation	Feistel	Feistel
Flexible	Yes	Yes	Yes	No
Features	Secure enough	Excellent Security	Adequate security Replacement for DES,	Not structure, Enough
Speed	fast	fast	Very slow	slow

Fig 1. Comparison of some of the encryption algorithms

IV. METHODOLODY

This research designs a frontend graphical user interface. It provides a user login. The user can chose the required encryption method. The file encryption module takes input of a filename and a key. And the file encryption is done. The public key is generated. Later the key and the same techniques are used to decrypt the file contents. The hybrid Algorithm uses RSA encryption to encrypt the private key of user and applies normal round encryption to the data with round specific keys generated from the user private key. The result is that user does not need to worry of the key. It can be shared but it can be decrypted only by the user to whom it was assigned. On decryption end the RSA encrypted key is to be provided. This file decryption must be done using the same mode as that of the encryption. In dynamic implementation, each user is assigned with public key. But the user has the d and n pair. Any file encrypted with user public key can only be decrypted by that user. For image cryptography, the data is converted into ASCII value and then 8 bit binary. Three pixels are read at a time in the image to be encrypted. The first eight RGB values are used to store one character which is eight bit binary. The module for comparison of the algorithms provides a better understanding of the algorithms.

Algorithm 1 The structure of RSA algorithm as follows.

1: **Input Values:** p and q
2: **Compute:**
3: n = p x q
4: (n) = (p-1) (q-1)
5: **Select Integer values:** e [(gcd (), e) - 1; 1 < e < ϕ (n)]
6: **Compute:** d de *mod* ϕ (n) = 1
7: C = Cg 1 *mod* (z)
8: **Encryption:** M < n C = M (*mod* n)
9: **Decryption:** CM = C(*mod* n)

Fig 2. RSA Algorithm

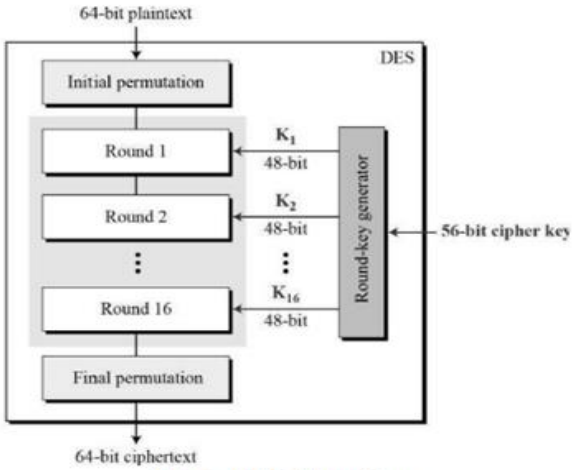


Fig 3. DES Algorithm

V. RESULTS

The file encryption module gives an application based technique for both encryption and decryption. On the other hand, RSA is one of the most secure encryption Algorithm, but using RSA algorithm for data encryption results in time consumption. Since, it is 10 times slower than DES. In this research, implementation is done to compare the effect asymmetric and symmetric encryption algorithm, considering the input size, time taken for execution and memory used. The image cryptography is also created using the pixel modification technique.

VI. CONCLUSION

In this research, the importance of encryption is network security is mentioned which helps in preventing several attacks in network. The encrypted data is more secure than the plain text. It protects data from attackers through various media. Different techniques can be incorporated based on the input type and the requirement.

VII. REFERENCES

[1]. Omar G. Abood, Shawkat K. Guirguis, A Survey on Cryptography Algorithms, International Journal of Scientific and Research Publications, Volume 8, Issue 7, July 2018 DOI: 10.29322/IJSRP.8.7.2018.p7978

[2]. Kajal, Rizwan Khan, Review on Network Security and Cryptography, International Journals of Advanced Research in Computer Science and Software Engineering ISSN: 2277-128X (Volume-8, Issue-6), June, 2018.

[3]. Kartik Sharma, Ashutosh Aggarwal, Tanay Singhania, Deepak Gupta, Ashish Khanna, Hiding Data in Images Using Cryptography and Deep Neural Network, Journal of Artificial Intelligence and Systems, 2019, 1, 143-162

[4]. Saurabh Sindhu, Divya Sindhu, Cryptographic Algorithms:Applications in Network Security, International Journal of New Innovations in Engineering and Technology, Volume 2, Issue 1, February 2017.

[5]. Ahmad Abusukhon, Muhammad Naveed Anwar, Zeyad Mohammad, BareeqAlghannam, A Hybrid Network Security Algorithm Based on Diffie Hellman and Text-to-Image Encryption Algorithm, Journal of Discrete Mathematical Sciences and Cryptography, 22 (1). pp. 65-81.ISSN 0972-0529, 2019

[6]. Dr. Sandeep Tayal, Dr.Nipin Gupta, Dr. Pankaj Gupta, Deepak Goyal, Monika Goyal, A Review paper on Network Security and Cryptography Algorithm, Advances in Computational Sciences and Technology, ISSN 0973-6107 Volume 10, Number 5 (2017) .

[7]. V.Esther Jyothi, Dr. BDCN Prasad, Dr Ramesh Kumar Mojjada, Analysis of Cryptography Encryption for Network Security, IOP Conf. Series: Materials Science and Engineering,

doi:10.1088/1757- 899X/981/2/022028, ICRAEM 2020.

[8]. Payal Kaushal, Navjot Kaur, A Review on Digital Image Watermarking, International Journal of Engineering Research & Technology, IJERT, December 2015.

[9]. Yahia Alemami, Mohamad Afendee Mohamed, Saleh Atiewi, Research on Various Cryptography Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S3, July 2019.

[10]. Fatma Mallouli, Aya Hellal, Nahla Sharief Saeed, Fatimah AbdulraheemAlzahrani, A Survey on Cryptography: comparative study between RSA vs ECC Algorithms, and RSA vs El-Gamal Algorithms, 6th IEEE International Conference on Cyber Security and Cloud Computing, 2019.

[11]. Rajani Devi. T, Importance of Cryptography in Network Security, International Conference on Communication Systems and Network Technologiesm, 2013.

[12]. Joselin.J, S.J.Brintha, V. Magesh Babu, Role of Digital Signature in Network Security and Cryptography, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015.

[13]. Abdalbasit Mohammed Qadir, NurhayatVarol, A Review Paper on Cryptography, IEEE, 2019.

[14]. Arvind K. Sharma, Dr.S.K. Mittal, Cryptography & Network Security Hash Function Applications, Attacks and Advances: A Review, International Conference on Inventive Systems and Control (ICISC 2019), IEEE.

[15]. Yong Huang, Fugui Chen, Peixin Qu, Research on Digital Signature Based on Digital Certificate, Proceedings of 14th Youth Conference on Communication.

[16]. Dr.Kamaljit I. Lakhtaria, Protecting Computer Network with Encryption Technique: A Study, International Journal of u- and e- Service, Science and Technology, Vol. 4, No. 2, June, 2011.

[17]. T. Resinger, I. Wagner, and E. A. Boiten, Security and Privacy in Unified Communication, ACm Computing Surveys, April, 2023.

[18]. S. Boumerdassi, A. Belmeguenai, D. Herbadji, and I. Haddad, Color image encryption based on an improved Fractional-order logistic map, International Journal of Electronic Security and Digital Forensics, 2023.

[19]. T. Dong and T. Huang, Neural Cryptography Based on Complex-Valued Neural Network, in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4999-5004, Nov. 2020, doi: 10.1109/TNNLS.2019.2955165.

[20]. SayantaniBasu, Marimuthu Karuppiah, MitaNasipuri, Anup Kumar Halder, Niranchana Radhakrishnan, Bio-inspired cryptosystem with DNA cryptography and neural networks, Journal of Systems Architecture, Volume 94, 2019, Pages 24-31, ISSN 1383-7621, https://doi.org/10.1016/j.sysarc.2019.02.005..

[21]. Mohammed Nazeh Abdul Wahid, Abdulrahman Ali, Babak Esparham, Mohammed Marwan, A Comparison of Cryptographic Algorithms:DES, 3DES, AES, RSA and Blowfish for Guessing Attacks Prevention, Journal of Computer Science Application and Information Technology, August, 2018.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Decentralized Network Server : Blockchain Technology and Future Regulations

Rakshith H R, Suriya R L, Rakshith C, Nageshwaran
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

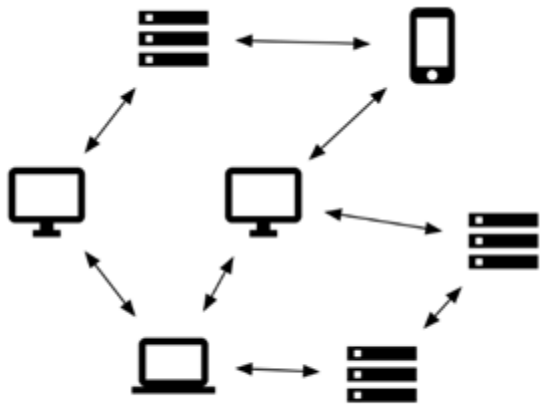
Blockchain, the foundation of Bitcoin, has received extensive attentions recently. Blockchain serves as an immutable ledger which allows transactions take place in a decentralized manner. Blockchain-based applications are springing up, covering numerous fields including financial services, reputation system and Internet of Things (IoT), and so on. However, there are still many challenges of blockchain technology such as scalability and security problems waiting to be overcome. This paper presents a comprehensive overview on blockchain technology. We provide an overview of blockchain architecture firstly and compare some typical consensus algorithms used in different blockchains. Furthermore, technical challenges and recent advances are briefly listed. We also lay out possible future trends for blockchain.
A decentralized network, anyone can participate and transact on the ledger. As a result, mechanisms must exist order to combat the vulnerabilities that arise from this design and to ensure that transactions are correct. Bitcoin, for example, is a decentralized blockchain that uses mining and proof-of-work to maintain the integrity of the ledger and to prevent people from corrupting the system.

Keywords: Blockchain, Decentralized server, Bitcoin and Technology

I. INTRODUCTION

In blockchain, decentralization refers to the transfer of control and decision-making from a centralized entity (individual, organization, or group thereof) to a distributed network. Decentralized networks strive to reduce the level of trust that participants must place in one another, and deter their ability to exert authority or control over one another in ways that degrade the functionality of the network.
Decentralization is actually not a new idea. When assembling a technological arrangement, three essential network structures are commonly thought of: centralized, decentralized, and disseminated. While blockchain technologies frequently utilize decentralized networks, a blockchain application itself simply can't be sorted as being decentralized or not. Maybe, decentralization is a sliding scale and should be applied to all parts of a blockchain application. By decentralizing the administration of and admittance to assets in an application, more prominent and more pleasant assistance can be accomplished. Decentralization typically has a few setoffs, for example, lower exchange. However, such

setoffs are worth it in comparison to the improved security and services they provide.



Decentralized

Blockchains are typically managed by a peer-to-peer network for use as a publicly distributed ledger, where nodes collectively adhere to a protocol to communicate and validate new blocks. Although blockchain records are not unalterable as forks are possible, blockchains may be considered secure by design and exemplify a distributed computing system with high Byzantine fault tolerance.

The blockchain was popularized by a person (or group of people) using the name Satoshi Nakamoto in 2008 to serve as the public transaction ledger of the cryptocurrency bitcoin, based on work by Stuart Haber, W. Scott Stornetta, and Dave Bayer. The identity of Satoshi Nakamoto remains unknown to date. The implementation of the blockchain within bitcoin made it the first digital currency to solve the double-spending problem without the need of a trusted authority or central server. The bitcoin design has inspired other applications and blockchains that are readable by the public and are widely used by cryptocurrencies. The blockchain is considered a type of payment rail.

II. TYPES OF BLOCKCHAIN

1. Public blockchain

How it works. The first type of blockchain technology is public blockchain. This is where cryptocurrency like Bitcoin originated and helped to popularize distributed ledger technology (DLT). It removes the problems that come with centralization, including less security and transparency. DLT doesn't store information in any one place, instead distributing it across a peer-to-peer network. Its decentralized nature requires some method for verifying the authenticity of data. That method is a consensus algorithm whereby participants in the blockchain reach agreement on the current state of the ledger. Proof of work (Pow) and proof of stake (PoS) are two common consensus methods.

2. Private blockchain

How it works. A blockchain network that works in a restrictive environment like a closed network, or that is under the control of a single entity, is a private blockchain. While it operates like a public blockchain network in the sense that it uses peer-to-peer connections and decentralization, this type of blockchain is on a much smaller scale. Instead of just anyone being able to join and provide computing power, private blockchains typically are operated on a small network inside a company or organization. They're also known as permissioned blockchains or enterprise blockchains.

3. Hybrid blockchain

How it works. Sometimes, organizations will want the best of both worlds, and they'll use hybrid blockchain, a type of blockchain technology that combines elements of both private and public blockchain. It lets organizations set up a private, permission-based system alongside a public permissionless system, allowing them to control who can access specific data

stored in the blockchain, and what data will be opened up publicly.

4. Consortium blockchain

How it works. The fourth type of blockchain, consortium blockchain, also known as a federated blockchain, is similar to a hybrid blockchain in that it has private and public blockchain features. But it's different in that multiple organizational members collaborate on a decentralized network. Essentially, a consortium blockchain is a private blockchain with limited access to a particular group, eliminating the risks that come with just one entity controlling the network on a private blockchain.

III. DECENTRALIZED BLOCKCHAIN SERVICE PROVIDER

There are not many true decentralized systems because value of the currency depends on the buying and selling of goods and services. The fundamental ideology of the cryptology to create a stateless transaction, freedom of expression without the interference of a centralized system and people to have self-sovereignty. which is achieved through this blockchain, But it can be monitor as the transactions happens mostly in the centralized market to buy and sell goods using crypto. As there are not many decentralized server markets.

One of the decentralized Cloud blockchain server is Sia

Sia is reinventing cloud storage. Sia's technology connects users who need file storage with hosts worldwide offering underutilized hard drive capacity. Blockchain technology secures their data and enables improved economics for users and hosts.

The Sia blockchain allows this marketplace to run without an intermediary. Sia secures storage transactions with smart contracts, creating a more reliable and affordable offering when compared to traditional cloud providers. No one person or organization can censor or deny access to data—not miners, not developers, nor any government. Sia's distributed, decentralized storage technology is stronger, more secure, more efficient, and more equitable for the entire ecosystem.

Decentralized main motive that neither individuals nor institutions should be beholden to large corporations. Sia was created to promote and safeguard users' freedoms. It's our attempt to bring decentralized storage built on blockchain technology to the mainstream without sacrificing price or performance.

Our long-term goal is for Sia to become the backbone storage layer of the Internet.

Nebulous, Inc. was established in 2014 and launched the Sia Storage Platform in 2015. Based in Boston, Nebulous is funded by Raptor Group, First Star Ventures, Fenbushi Capital and INBlockchain.

IV. WORKING OF DECENTRALIZED PEER TO PEER NETWORK WITH BLOCKCHAIN TECHNOLOGY

Files Are Divided Prior To Upload

The software divides files into 30 segments before uploading, each targeted for distribution to hosts across the world. This distribution assures that no one host represents a single point of failure and reinforces overall network uptime and redundancy.

File segments are created using a technology called Reed-Solomon erasure coding, commonly used in CDs and DVDs. Erasure coding allows Sia to divide files in a redundant manner, where any 10 of 30 segments can fully recover a user's files.

This means that if 20 out of 30 hosts go offline, a can user is still able to download her files.

Each File Segment Is Encrypted

Before leaving a renter's computer, each file segment is encrypted. This ensures that hosts only store encrypted segments of user data.

This differs from traditional cloud storage providers like Amazon, who do not encrypt user data by default. Sia is more secure than existing solutions because hosts only store encrypted file segments, rather than whole files. Uses the Threefish algorithm, an open source, secure, high-performance encryption standard.

Files Are Sent to Hosts Using Smart Contracts

Using the blockchain, renters form file contracts with hosts. These contracts set pricing, uptime commitments, and other aspects of the relationship between the renters and the hosts.

File contracts are a type of smart contract. They allow us to create cryptographic service level agreements (SLAs) that are stored on the Sia blockchain.

Since file contracts are automatically enforced by the network, has no need for intermediaries or trusted third parties.

Storage proofs can be made possible by a technology called Merkle trees. Merkle trees make it possible to prove that a small segment of data is part of a larger file. The advantage of these proofs is that they are very small, no matter how large the file is. This is important because the proofs are stored permanently on the blockchain.

Table 1: Difference between Centralized, decentralized and distributed server.

	Centralized	Distributed	Decentralized
Network/hardware resources	Maintained & controlled by single entity in a centralized location	Spread across multiple data centers & geographies; owned by network provider	Resources are owned & shared by network members; difficult to maintain since no one owns it
Solution components	Maintained & controlled by central entity	Maintained & controlled by solution provider	Each member has exact same copy of distributed ledger
Data	Maintained & controlled by central entity	Typically owned & managed by customer	Only added through group consensus
Control	Controlled by central entity	Typically, a shared responsibility between network provider, solution provider & customer	No one owns the data & everyone owns the data
Single Point of Failure	Yes	No	No
Fault tolerance	Low	High	Extremely high
Security	Maintained & controlled by central entity	Typically, a shared responsibility between network provider, solution provider & customer	Increases as # of network members increase
Performance	Maintained & controlled by central entity	Increases as network/hardware resources scale up and out	Decreases as # of network members increase
Example	ERP system	Cloud computing	Blockchain

V. CONCLUSION

Decentralized blockchain does not have a single point of failure. It provides more security and privacy for the user data to carry of the action. If the decentralized storage is provided the user will be able be create his own website without depending upon big centralized networks as they can have a separate cryptocurrency for their website or market for transaction so dependence on one crypto currency can be avoided. And the future of the crypto will be based on is fully decentralized blockchain and ledger usage. Decentralization provides high Proof-of-work: a function to deter denial of service attacks and other potential spam on a network by requiring some level of work that is usually costly and time consuming. In the Bitcoin network, miners must complete a proof-

of-work in order to regulate the rate at which new blocks are generated. Immutability in decentralized server implies that data cannot be edited or deleted by participants or system administrator after it is written to the ledger.

VI.REFERENCES

[1]. The Basics of Bitcoins and Blockchains by Antony Lewis
[2]. Mastering Blockchain by Lorne Lantz, Daniel Cawrey
[3]. 5 Blockchain and the Law: The Rule of Code by Aaron Wright & Primavera De Filippi
[4]. Researchgate.net
[5]. Degruyter.com

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Customer Behavioral Segmentation

Sanidhya Singh, Chempavathy B, Krishna Agrahari
Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Customer segmentation is the practice of dividing a company's customers into groups that reflect the similarity between customers in each group. The goal of customer segmentation is to decide how to interact with customers in each segment to maximize each customer's value to the business. Customer segmentation has the potential to allow marketers to target every customer in the most effective way. Using the vast amount of customer (and potential customer) data available, a customer segmentation analysis allows marketers to identify discrete groups of customers with a high degree of accuracy based on demographic indicators. ,behavioral and others. Since the marketer's goal is often to maximize the value (revenue and/or profit) of each customer, it is essential to know in advance how a particular marketing action will influence the customer. Ideally, such "action-oriented" customer segmentation will not focus on the short-term value of a marketing action, but on the long-term impact of the customer lifetime value (CLV) that the marketing action will have. It is therefore necessary to group or segment customers according to their CLV.

Keywords- Demographic Segmentation, Behavioral Segmentation, Needs-based Segmentation, Value-based Segmentation

I. INTRODUCTION

Customer segmentation is the process by which you divide your customers based on common characteristics, such as demographics or behaviors, so that you can market to those customers more effectively. These customer segment groups can also be used to start discussions about creating a marketing personality. In fact, customer segmentation is often used to inform messaging, brand positioning, and improve the way a business sells. Therefore, marketing personalities must be closely aligned with these customer segments to be effective.

The marketing "persona" is, by definition, a personification of a customer segment, and it is not uncommon for companies to create multiple personas to match their different customer segments. But for that to happen, a business needs a strong set of customer segments to build on. Which brings us to the next section, distinguishing the difference between customer segmentation and market segmentation, so that your segmentation is as accurate as possible?

Compared to customer segmentation, market segmentation is more general and analyzes the entire market. While market segmentation is about the

entire market, customer segmentation is your part of the market. For example, if you are in the vehicle sales industry and typically sell to businesses, your customer segment is B2B and you can compare customers who are likely to buy large commercial trucks with small business vans. These two customers have different needs, and depending on the correlation you find, they could become two different customer segments to focus on. However, if you segment the entire market, you can compare people looking for a minivan to a sports car. which is much larger. In this case, most producers in the market will not serve the entire market, so it is more efficient to focus on the selling element. You will get a better reward if you target one or two targeted customer segments, rather than the entire market.

II. LITERATURE SURVEY

1. Determine your customer segmentation goals.

Think about why you're creating a customer segmentation strategy: ask yourself why you spend time segmenting and what you hope to get out of the process.

To do this, check out the list of common reasons companies choose to segment customers that we looked at earlier. Determine the results you are looking to achieve so that you can develop the rest of your strategy in a way that helps you achieve them.

One thing to keep in mind when developing your customer targeting goals is that they will be unique to your business: customer targeting is not a one-size-fits-all process.

For example, your goal count will be unique to you based on the size, type, and industry of your business, as well as your customers. Additionally, your goals may be relevant to multiple teams (eg, marketing, sales, service) or to a particular department. Therefore, when determining your customer segmentation goals, use your company's specific business characteristics and needs as a launching pad.

More information about this source text The source text is required for additional translation information post comments side panels

2. Segment your customers into groups of your choice.

Once you have an idea of what you're looking to get out of the customer segmentation process, decide how you're going to segment your customers. See Models and types of customer segmentation to determine how you will do this. There is no right or wrong answer here – it depends on your business, your customers, and the goals you set in the previous step.

For example, if you want to share targeted ads with West Coast audience members and customers in hopes of increasing conversions in that region, you can geo-target your customers.

3. Target and reach your customer segments.

Once you've segmented your customers, it's time to figure out how you'll target them in your organization. By ensuring that all departments (eg, marketing, sales, and service) understand how their customers are segmented, members of these departments will be able to effectively target their customers with their work.

For example, think back to our example above about targeting ads to your West Coast customers:

Marketing can tailor and personalize content to engage, educate, and meet the needs of your West Coast audience members to generate leads and brand awareness.

Sales can identify the common traits shared by your most qualified West Coast customers, as well as the best ways to contact and communicate with them, to increase conversions.

The service can use its customer segments to prepare materials and resources for them based on the

challenges West Coast customers are most likely to face.

4. Run customer segmentation analysis.

Analyzing your segmentation efforts will give you insight into how you've organized your customers, that way you can[3] make updates and changes as needed. Check in with your marketing, sales, and service teams (as often as you'd like) to get their input on needed targeting adjustments. You can also experiment with new ways to group your customers to decide what makes the most sense.[10]

You can also collect feedback from your customers to more effectively segment them into appropriate groups. For example, you can conduct surveys to improve your behavioral targeting by asking customers about feature usage and product usage habits/trends. This will allow you to target customers more precisely based on their specific behaviors. Finally, consider how your customers are segmented each time you update your product or service, change brands, or revise your buyer personas; these changes may require the reorganization of some or all of its customer segments.

Types Of Customer Segmentation:

Importance of customer segmentation:

Companies not only strive to divide their customers into measurable segments based on their needs, behaviors or demographics, but also seek to determine the profit potential of each segment by analyzing its impacts on revenue and costs. Value-based segmentation evaluates customer groups in terms of the revenue they generate and the costs of establishing and maintaining relationships with them. It also helps companies determine which segments are the most and least profitable so they can adjust their marketing budgets accordingly.

Customer segmentation can have a great effect on customer management in that, by dividing customers into different groups that share similar needs, the company can market to each group differently and focus on what each kind of customer needs at any given moment. Large or small, niche customer segments can be targeted depending on the company's resources or needs.[5]

In B2B marketing, companies are concerned with the job titles of decision makers, the industry, whether the company is public or private, its size, location, shopping habits and technology at its disposal, for example.In B2C marketing, companies are concerned with particular customers' profiles, attitudes and lifestyles. B2C companies may also be concerned with geographic location. B2C companies who segment customers based on their geographic location can tailor offers based on regional events and preferences. B2C companies can also customize offers based on the predominant languages spoken in each region.

B2B customer segmentation approaches include vertical or horizontal alignments. In vertical segmentation, companies select certain industries or job titles that are likely to find their products appealing, and then focus their marketing efforts on the segments they deem most ready to buy. The advantage of vertical segmentation is that companies can offer services tailored to particular industries. The needs of the financial services sector are different from those of the health sector. If each segment were

offered services tailored to that industry, adoption and satisfaction could increase.

In horizontal segmentation, companies simply focus on one job title across a wide range of industries and organizations. The advantage of horizontal segmentation is to put more emphasis on the needs of particular positions or functions. For example, the focus on CFOs can create product materials, website postings, and email newsletters specifically tailored to this role.

Customer segmentation vs. market segmentation:

Companies can use marketing automation software to define and create customer segments. Customer segments can be based on demographic, psychographic, and activity-based data, such as actions users have taken on a website. Companies use marketing automation software to set up, plan, and execute campaigns for particular customer segments.

Customer segmentation is different from market segmentation. An example of market segmentation is grouping customers based on the products or services they buy. A company can perform market segmentation based on different industries, such as software, professional services, and training. The company can then allocate resources to each market segment and use separate marketing and advertising activities for each.

III. CONCLUSION

This document provides a step-by-step process for identifying, prioritizing, and targeting your best customer segments. Simply following it does not guarantee success. To be effective, you need to prepare and plan for the different challenges and obstacles each step can present, and always make sure to adapt your process to any new information or feedback that may change your outcome. Also, you cannot force the power of this process on your company. If the key stakeholders who will be affected by the current best customer segmentation process do not fully accept it, the resulting results will be relatively insignificant. However, if you properly manage today's best customer segmentation process, the impact it can have on every part of your organization: sales, marketing, product development, customer service, and more. - it's huge. Your business will be more customer-centric and clearer on the market, allowing you to scale much more predictably and efficiently. Ultimately, this means you no longer need to support all customers who are willing to pay for your product or service, allowing you to focus on a specific subset of customers that present profitable opportunities and efficient use of resources. This is essential for all businesses, of course, but in the expansion stage it can often be the difference between incredible success and surefire failure.

IV.REFERENCES

[1]. Dr. R. Gardener "The Essential R Reference' (2014),

[2]. Concepts of customer segmentation httxr//www.business-science.io

[3]. https://labs.openviewpartners.com/custorner-segmentation/

[4]. Source data related to our analysis has been collec-tectfromhttps://github-com/mdancho84/orcl erSimulatoR/tree/master/data
[5]. https://www.kaggle.com/
[6]. https://wvvwx-project.org/
[7]. https://www.rstudio.com/
[8]. Alexander, C.E., Wilson, CA. & Foley, D.H. 2005. Agricultural input market segments: who is buying what? Journal of Agribusiness, 23(21:113-132. [Online] Available from: htto://0-
[9]. www.jab.uga.edu.innopac.up.sc.za/Library/F05-01.pdf [Accessed: 2013-01-31]. Babble, E. 2011. The basics of social research. 5th ed. Belmont, USA: Wadsworth. Bailey, C., Baines, P.R., Wilson, H. & Clark, M. 2009. Segmentation and customer insight in contemporary services marketing practice: why grouping customers is no longer enough, Journal of Marketing Management, 25(3):227- 252. [Online]
[10]. Available from: Emerald: http://dx.dolorg/10.1362/0267257 09X429737 [Accessed: 2012-11-10]. Baker, MJ. 2002.
[11]. Sampling. The Marketing Review, 3(1):103-120. [Online] Available from: gatratlen: Business Sourcehtto://0- web.ebscohost-com.innopac.up.ac.za/ehost/detai 1?vid=3&hid=12213ai6=127f865 1-ed20-48f3-9298-753.812c57bd9640sessionmerl15&IxlatamilnNpOG inzyytiyacttbG177s77Y29yeZT1zaJCRPtdbinbuh &AN =8561083 [Accessed: 2012-08-101.2

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Evaluation of Cyber Attacks

Pragya Goel, Pragati Subhadarshinee
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The possibility of Cybercrime isn't new, yet there is huge disarray among scholastics, PC security specialists and clients regarding the degree of genuine Cybercrime.[1] In this paper, we investigate the broadness of PC based wrongdoing, giving a meaning of the arising terms "Cybercrime" and "crimeware". We then partition Cybercrime into two unmistakable classes: Type I Cybercrime, which is for the most part innovative in nature, and Type II Cybercrime, which has a more articulated human component. We then, at that point, utilize two contextual analyses to show the job of crimeware in various kinds of Cybercrime, and express a few impressions on the job of comprehension during the time spent Cybercrime. Such goes after having been expanding in number and intricacy throughout recent years.

Keywords—Cybercrime, Crimeware, Phishing, DDoS, Cryptomining.Section 1 tells about what is cybercrime, what are the ways to prevent it and how these hackers attack our systems.

I. INTRODUCTION

Many individuals actually have no idea as what to do in the event that they are being cheated disconnected or on the web. A many individuals record grumbles every day on the public authority approved site, yet what are the consequences of those gripes. Certain individuals who we can say luck out may accept their cash back or the data and so on. Yet, this is just 2% of the complete gripes enlisted what might be said about 98% individuals who are still in this trust that something will improve or there is an answer for this multitude of issues.

Conversations of Cybercrime can be found in assorted sources including scholarly diaries, generalist PC magazines, paper articles and on the web; it has been the subject of films, TV projects and radio communicates. Be that as it may, in spite of an evident acknowledgment of and knowledge of the term, there exist emphatically differed perspectives on what Cybercrime is. This absence of definitional clearness is hazardous as it impacts each feature of counteraction and remediation. Moreover, research shows that the quantity of individuals and organizations affected by different kinds of seen cybercrime is developing with no indications of declining.

Notwithstanding cybercrime, digital assaults can likewise be related with digital fighting or cyberterrorism, like hacktivists. Inspirations can fluctuate, at the end of the day. Also, in these inspirations, there are three principle classes: criminal, political and individual.

Criminally spurred assailants look for monetary profit through cash robbery, information burglary or business interruption. In like manner, the by and by spurred, for example, disappointed current or previous representatives, will take cash, information or a simple opportunity to disturb an organization's framework. Be that as it may, they basically look for requital. Socio-political inspired aggressors look for consideration for their causes. Subsequently, they spread the word about their assaults for the public-otherwise called hacktivism.

Other digital assault inspirations incorporate reconnaissance, spying-to acquire an unjustifiable benefit over contenders and scholarly test.

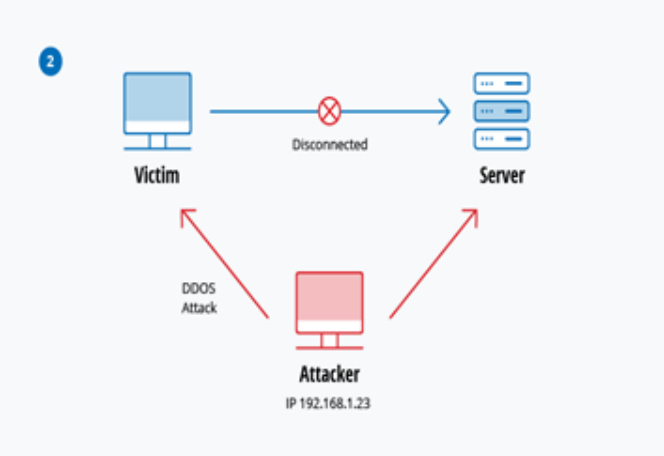


Fig.1. Cyber Attack[7]

Section 2 deals with how Cyber Crimes are practiced and performed. Section 3 talks about the concept of hacking and its types. Finally, Section 4 concludes the paper by dealing with a brief note on how can we prevent cyber-attacks and how it is useful to us.

II. CYBER CRIMES

Numerous digital assaults are artful, with programmers spotting weaknesses in a PC framework's guards and taking advantage of them. This might include observing blemishes in the code of a site, that permits them to embed their own code and afterward sidestep security or validation processes. It could likewise mean they introduce 'malware' - programming which is explicitly intended to harm a framework - through a weak outsider site.

In spite of the fact that phrasing, for example, 'digital assaults' and 'programmers' may invoke pictures of complex groups of PC specialists with cutting edge gear, poring over lines of code, the fact of the matter is frequently very unique. Digital assaults are substantially more prone to happen through unremarkable blunders like a client picking a simple to-figure secret word or not changing the default secret word on something like a switch. Some of the techniques are as follows:

A. PHISHING

Phishing[2] is a procedure utilized by cheats in which they mask themselves as dependable substances and they accumulate the objective's touchy data, for example, username, secret key, and so on, Phishing is a method for getting individual data using deceiving messages and sites. Phishing endeavors to convince an email beneficiary that the message is something that they need, like a solicitation from their bank or a letter from a colleague, and that they ought to click a connection or download a connection.

III. TYPES OF PHISHING ATTACKS

- Misleading phishing: Sending a bogus email as a group with a source of inspiration that requires the beneficiary to tap on a connection is known as tricky phishing.
- DNS-based phishing: Phishing that compromises the uprightness of the area name look-into process is known as DNS-based phishing.

The following are instances of DNS-based phishing:

- Documenting of harming reports by has
- Tainting the DNS store of the client
- Compromising the intermediary server

- Smishing: This kind of phishing is a variety of email-based phishing tricks. As clients develop more overpowered by steady messages and more dubious of spam, instant messages have turned into a more appealing assault vector, taking advantage of the closer connection that individuals have with their telephones. Consequently, programmers, nowadays, are bound to embrace smishing.
- Skewer phishing: Spear phishing is a social designing procedure. It is a customized phishing assault that objectives a particular individual, association, or business. Cybercriminals utilizing lance phishing plan to take privileged data about an association, for example, login accreditations, or introduce malware in the association.
- Whaling: In this kind of phishing, assailants target senior leaders of an organization or other high-profile targets. The main role of aggressors is to persuade a casualty to move a gigantic measure of cash or uncover some touchy data.
- Vishing:Vishing[3], otherwise called voice phishing includes a vindictive guest who professes to counterfeit personalities, for example, being technical support, government specialist, and so on and removes individual data, for example, bank or Mastercard subtleties. This is perhaps the most common sort of phishing and it frequently works out, wind up tricking many individuals consistently.

IV. PHISHING ATTACK PREVENTION

- Clients should be constant and wary. Minor blunders in a noxious email can uncover the shipper's authentic personality. It very well may be linguistic mistakes or an adjustment of an area name. Clients should not tap on irregular messages without first checking the source out.
- Associations ought to play it safe to stay away from phishing tricks. Two-factor validation, otherwise called 2FA, is the most reasonable answer for ruining phishing endeavors since it adds an additional a layer of confirmation and check while interfacing with endorsed programming or applications. With each sign in endeavor, the most pervasive 2FA frameworks forestall phishing assaults by creating another one-time code.
- Authorizing secure practices, for example, not tapping on outer email joins and instructive endeavors, can assist with diminishing the danger of phishing endeavors.

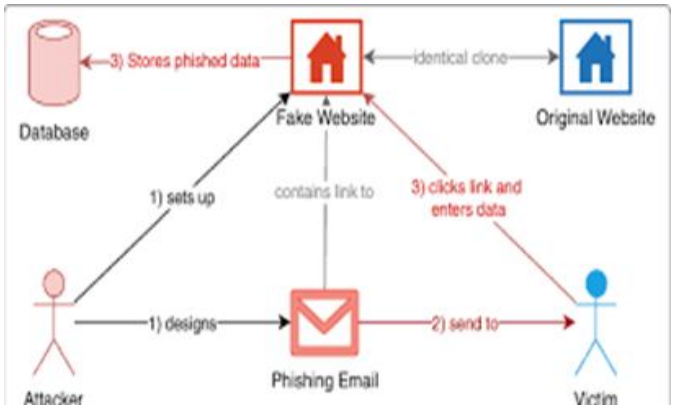


Fig.2. Email Hacking Using Phishing[8]

A. DISTRIBUTED DENIAL OF SERVICE

A conveyed disavowal of-administration (DDoS) assault is a malevolent endeavor to disturb the ordinary traffic of a designated server, administration or organization by overpowering the objective or its encompassing framework with a surge of Internet traffic.

DDoS assaults accomplish adequacy by using various compromised PC frameworks as wellsprings of assault traffic. Taken advantage of machines can incorporate PCs and other arranged assets like IoT gadgets?

From a significant level, a DDoS assault resembles a surprising gridlock stopping up the parkway, keeping standard traffic from showing up at its objective.

V. IDENTIFYING DDOS ATTACKS

The clearest side effect of a DDoS assault is a site or administration unexpectedly turning out to be slow or inaccessible. Be that as it may, since various causes — such an authentic spike in rush hour gridlock — can make comparable execution issues, further examination is normally required. Traffic investigation apparatuses can assist you with detecting a portion of these indications of a DDoS assault [4]. Dubious measures of traffic beginning from a solitary IP address or IP range. A surge of traffic from clients who share a solitary conduct profile, for example, gadget type, geolocation, or internet browser adaptation

An unexplained flood in solicitations to a solitary page or endpoint. Odd traffic examples, for example, spikes at odd hours of the day or examples that have all the earmarks of being unnatural (for example a spike like clockwork)

VI. PREVENTION OF DDoS ATTACKS

1. Know your organization's traffic

Each association's framework has regular Internet traffic designs — know yours. At the point when you comprehend your association's typical traffic design, you'll have a standard. Like that, when uncommon action happens, you can recognize the side effects of a DDoS assault.

2. Make a Denial of Service Response Plan

Do you have any idea what will happen when and assuming that a DDoS assault occurs? How might your association answer? By characterizing an arrangement ahead of time, you'll have the option to answer rapidly and proficiently when your organization is designated.

This can take some preparation; the more complicated your foundation, the more nitty gritty your DDoS reaction plan will be. No matter what your organization's size, notwithstanding, your arrangement ought to incorporate the accompanying:

- A frameworks agenda
- An automatic reflex group
- Obvious warning and acceleration methodology.
- A rundown of inward and outer contacts that ought to be educated about the assault
- A correspondence plan for any remaining partners, similar to clients, or merchants

3. Make your organization strong

Your framework ought to be all around as versatile as conceivable against DDoS assaults. That implies more than firewalls since some DDoS assaults target firewalls. Rather think about ensuring you're not keeping every one of your eggs in a similar container — put server farms on various organizations, ensure that not every one of your server farms are in similar actual area, put servers in various server farms, and be certain that there aren't spots where traffic bottlenecks in your organization.

4. Practice great digital cleanliness

It's implied that your clients ought to be participating in best security works on, including evolving passwords, secure verification works on, knowing to keep away from phishing assaults, etc. The less client blunder your association illustrates, the more secure you'll be, regardless of whether there's an assault.

5. Increase your transmission capacity

Assuming DDoS is making a gridlock in your organization, one method for making that gridlock less serious is to enlarge the interstate. By adding more data transfer capacity, your association will actually want to retain more to assimilate a bigger

volume of traffic. This arrangement won't shut down all DDoS assaults, in any case. The size of volumetric DDoS assaults is expanding; in 2018, for instance, a DDoS assault bested 1 Tbps in size interestingly. That was a record, until a couple of days after the fact, when a 1.7 Tbps assault happened.

6. Exploit hostile to DDoS equipment and programming

DDoS assaults have been around for some time and a few sorts of assaults are exceptionally normal. There are a lot of items that are ready to repulse or relieve specific convention and application assaults, for instance. Exploit those instruments.

7. Move to the cloud

While this will not take out DDoS assaults, moving to the cloud can alleviate assaults. The cloud has more transmission capacity than on-premise assets, for instance, and the idea of the cloud implies numerous servers are not situated in a similar spot.

8. Know the side effects of an assault

Your organization dials back mysteriously. The site closes down. Out of nowhere, you're getting a ton of spam. These can all be indications of a DDoS assault. Provided that this is true, the association ought to examine.

9. Re-appropriate your DDoS assurance

A few organizations offer DDoS-as-a-Service. A portion of these organizations work in scaling assets to answer an assault, others support safeguards, yet, others alleviate the harm of a continuous assault.

10. Screen for uncommon action

When you know your run of the mill movement and the indications of an assault, screen your organization for odd traffic. By observing traffic progressively, your association will actually want to recognize a DDoS assault when it begins and moderate it.

VII. TYPES OF DDOS ATTACKS

1. APPLICATION LAYER ATTACK

At times alluded to as a layer 7 DDoS assault (regarding the seventh layer of the OSI model), the objective of these assaults is to deplete the objective's assets to make a refusal of-administration.

The assaults focus on the layer where site pages are created on the server and conveyed because of HTTP demands. A solitary HTTP demand is computationally modest to execute on the client side, however it tends to be costly for the objective server to answer, as the server frequently stacks different records and runs data set questions to make a site page.

Layer 7 assaults are challenging to guard against, since separating malignant traffic from real traffic can be hard.

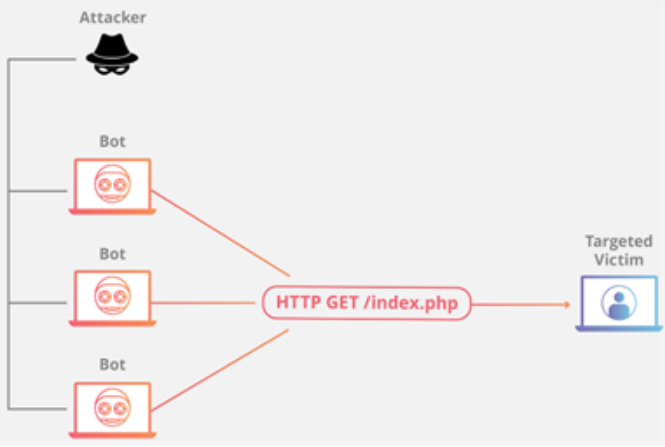


Fig. 3. Example of application layer attack[9]

VIII. HOW TO PREVENT THESE APPLICATION LAYER ATTACKS

Since DDoS assaults can be mind boggling in nature and a decided assailant will quickly change the assault vector to keep away from alleviation, the IDMS ought

to utilize a bunch of strategies to investigate and obstruct these sorts of DDoS assaults.

Best practices to guard against continually developing sorts of disavowal of administration assaults include:

- Use stream telemetry examination enhanced with conduct investigation to distinguish irregularities and assaults. Zero in on it is ordinary to grasp what. This will work on distinguishing proof of anomalies.
- Utilize an IDMS to recognize strange way of behaving and application layer goes after that require progressed and dynamic alleviation; and involving this methodologyrelated to BGP Flow Spec Offload when and where proper.

Whenever carried out effectively, these DDoS security methods will drive the assailant to act like typical clients, delivering the DDoS assault ineffectual and considering the utilization of use level investigation to distinguish any strange traffic or use designs.

PROTOCOL ATTACKS

Line Gateway Protocol (BGP) commandeering is an incredible illustration of a convention that can turn into the premise of a DDoS assault. BGP is utilized by network administrators to declare to different organizations how their location space in designed. Assuming a troublemaker figures out how to send a BGP update that is dared to be credible then traffic expected for one organization can be steered to an alternate organization and the false traffic can cause asset consumption and clog. Since BGP is utilized by a huge number of organization administrators all over the planet, a move up to a safer adaptation of the convention would be both confounded and extravagant to send. Other convention assault models incorporate SYN flood and Ping of Death.

Section 3 deals with the concept of hacking, what is it and what are the various methods of hacking.

HACKING

Hacker is the person who is savvy to the point of taking the data, for example, business information, individual information, monetary data, Visa data, username and secret word from the framework he is unapproved to get this data by assuming unapproved command of that framework utilizing various instruments and strategies.[5]

These weakness or shortcomings can be utilized by the programmer to perpetrate wrongdoing or to take benefits of remote machine.

TYPES OF HACKERS

1. WHITE HAT HACKERS

A white cap (or a white cap programmer) is a moral security programmer. Moral hacking is a term intended to infer a more extensive classification than just entrance testing. Under the proprietor's assent, white cap programmers expect to distinguish any weaknesses the ongoing framework has. Appeared differently in relation to the dark cap, a vindictive programmer, the name comes from Western movies, where courageous and opposing cowpokes could customarily wear a white and a dark cap, separately. There is a third sort of programmer known as a dim cap who hacks with honest goals yet now and again without authorization.

White cap programmers may likewise work in groups called "shoes and additionally programmer clubs", red groups, or tiger groups.

2. BLACK HAT HACKERS

Blackhat programmers, frequently known basically by the term blackhats (think rancher films), are the trouble makers of the programmer world. Such programmers frequently have no specific consideration for law and order, the frameworks that they upset, for sure sick impacts that they cause.

Blackhats are recognized from whitehats, the heroes, who are many times tracked down attempting to thwart the endeavors of the blackhats, and grayhats, who ride the line between the two, frequently crossing from one side to the next.

Recognizing an assailant as a blackhat frequently infers that they have a specific degree of ability at going after and taking advantage of frameworks and organizations, in abundance of the typical content youngster. Blackhats might go after a framework or organization in light of an assortment of inspirations.

3. GREY HAT HACKERS

A dark cap programmer (likewise spelled dim cap programmer) is somebody who might abuse moral guidelines or standards, however without the malevolent goal credited to dark cap programmers. Dark cap programmers might participate in rehearses that appear to be not exactly totally above board, yet are frequently working for a long term benefit. Dark cap programmers address the center ground between white cap programmers, who work in the interest of those keeping up with secure frameworks, and dark cap programmers who act vindictively to take advantage of weaknesses in frameworks.

ETHICAL HACKING

Ethical Hacking[6] includes being used of hacking instruments, stunts and strategies to distinguish weaknesses to guarantee framework security. It is likewise called as infiltration testing.

1. ETHICAL HACKER

A moral programmer is a master's in network safety and who goes after a security framework for its proprietors and looking for weaknesses that a noxious programmer could take advantage of.

SKILLS REQUIRED

1. Moral Hacker has top to bottom information on pretty much all working frameworks, including all famous, broadly utilized working frameworks like Windows, Linux, Unix, and Macintosh.
2. These moral programmers are exceptionally talented at systems administration and investigating abilities of equipment and programming.
3. They probably definite information about more established, progressed, complex assaults.

DEVICES VULNERABLE TO HACKING:

Savvy Devices:Savvy gadgets, for example, cell phones, are worthwhile focuses for programmers. Android gadgets, specifically, have a more open-source and conflicting programming advancement process than Apple gadgets, which seriously endangers them of information robbery or debasement. Be that as it may, programmers are progressively focusing on the large numbers of gadgets associated with the Internet of Things (IoT).

Webcams:Webcams incorporated into PCs are a typical hacking objective, primarily in light of the fact that hacking them is a basic cycle. Programmers regularly get to a PC utilizing a Remote Access Trojan (RAT) in rootkit malware, which permits them to keep an eye on clients as well as perused their messages, see their perusing action, take screen captures, and seize their webcam.

Switches:Hacking switches empowers an aggressor to get close enough to information sent and got across them and organizations that are gotten to on them. Programmers can likewise capture a switch to complete more extensive malevolent demonstrations like dispersed forswearing of-administration (DDoS) assaults, Domain Name System (DNS) satirizing, or cryptomining.Email:Email is perhaps the most well-known focus of cyberattacks. It is utilized to spread malware and ransomware and as a strategy for phishing assaults, which empower aggressors to target casualties with malevolent connections or connections.

Jailbroken Phones: Jailbreaking a telephone implies eliminating limitations forced on its working framework to empower the client to introduce applications or other programming not accessible through its true application store. Beside being an infringement of the end-client's permit concurrence with the telephone engineer, jailbreaking uncovered numerous weaknesses. Programmers can target jailbroken telephones, which permits them to take any information on the gadget yet in addition stretch out their assault to associated organizations and frameworks.

IX. CONCLUSION

ICT frameworks have turned into a spine of present-day power networks. Digital security is significant for soundness and dependability of the brilliant matrix. This paper is a best-in-class review of network protection R&D for a savvy framework.

Weaknesses are progressively present in the digital power framework climate because of the developing reliance on PC frameworks and computerized correspondence. Since there are impediments for firewalls to distinguish pernicious parcels, ADSs/IDSs are basic to identify oddities inside a private organization (e.g., LAN, HAN, and NAN). Besides, the execution of discovery frameworks ought to meet the prerequisites forpower frameworks, for example, exactness and correspondence delay. With apractical CPS testbed, analysts can test their digital security frameworksto assess whether necessities are met.Cyber attacks can be reduced by:Training or spreading awareness among people, keeping systems and software up to date, ensure end to end encryption, always back up data, install firewalls and wifisecurity should be ensured.

X. REFERENCES

[1]. Chawki, M.: Cybercrime in France: an overview. Computer Crime Research Center. December, 2005.

[2]. S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor and J. Downs, "Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions", Proceedings of the 28th international conference on Human factors in computing systems ser. CHI'10. New York NY USA:ACM, pp. 373-382, 2010.

[3]. R. Prasad, "Handheld Devices," in Anesthesia Informatics: Health Informatics, Springer, 2009, pp. 409-424.

[4]. Agrawal, S., Das, M.L.: Internet of Things – A Paradigm Shift of Future Internet Applications, Institute of Technology, Nirma University, Ahmedabad 382 481, 08-10 (2011).

[5]. IBM Research Division, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA

[6]. Akhgar, S., Yates, B. 2013. Strategic Intelligence Management, 1st Edition, Butterworth-Heinemann, 9780124071919, 56-255

[7]. Joel L. Fernandes, Ivo C. Lopes, Joel J. P. C.Rodrigues, Sana Ullah, Performance Evaluation of RESTful Web Services and AMQP Protocol, Fifth International Conference on Ubiquitous and Future Networks (ICUFN), 2-5 July 2013, pp. 810-815.

[8]. B. Parno, C. Kuo, and A. Perrig, “Phoolproof phishing prevention,” in International conference on financial cryptography and data security, 2006, pp. 1–19.

[9]. D. Andersen, H. Balakrishnan, M. Frans Kaashoek, and R. Morris. Resilient overlay networks. In Proceedings of 18th ACM Symposium on Operating Systems Principles, Oct. 2001.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Facility Layout Design Using Genetic Algorithm

Tejashwin U, Syed Mohammed Sana, Indrani C, Sushruth AC
Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

In this paper, we present a heuristic for designing facility layouts that are convenient for scheming a unidirectional loop for material handling. We use a genetic algorithm where the objective function, crossover and mutation operators have all been designed specifically for this purpose. The design is made under a flexible bay structure; comparisons are made with other layouts from the literature that were designed under the flexible bay structure.

Keywords—Facilities layout problems, Quadratic assignment problems, Genetic algorithms.

I. INTRODUCTION

The facility design problem is concerned with the conceptual design of the physical enclosure of manufacturing or service systems. It is crucial to integrate the designing principles of the facility layout and material flow pattern. The four principal and interdependent decisions in the conceptual design of the physical enclosure of manufacturing or service facilities are: 1) The determination of the size and shape of the facility; 2) The arrangement of the work centers as a set of fully packed right-angle polygons represented as a block layout; 3) The design of a connected flow network on the contour lines defined as the boundaries of the work centers; 4) The determination of the number and locations of P and D stations on the counter-lines of the flow network covering each work center. Although these problems are closely related, traditionally, they have been solved separately in a sequential manner because of the computational intractability of the integrated design problem. The inputs of a facility layout design problem are the number of work centers, the area required for each work center, the material flow between each pair of work centers, and shape constraints for the work centers. The work centers must be placed inside a practically rectangular building such that all constraints are met and the material handling cost is minimized. The flexible bay structure was first introduced in equation set 2. In the flexible bay structure (FBS), all the work centers are placed inside parallel bays. Equation 3 to 5 are examples of researches that designed the layout based on FBS. In this paper, we adopt this basis and concentrate on creating a layout that will be equipped with a loop material flow pattern

II. LITERATURE

The approximate solution approaches in meta-heuristics do not rely on the analytical functions and features of the problem which makes them a compatible solution approach to combinatorial optimization problems. The genetic algorithm solution methodology has attracted a number of users in the recent era to expedite the search for optimal solution in the combinatorial problems. In recent years, genetic algorithm has been approved as an innovative search approach used to solve the facility layout problems by many researchers (Banerjee and Zhou, 1995; Azadivar and Wang, 2000; Wu and Appleton, 2002; and Wang et al., 2005). Suresh and Sahu (1995) stated the problem of irregular shapes of departments could be transformed into a unit area of department by allocating a sufficient number of unit squares to satisfy the area requirements. A crossover operator has been developed to ensure the diversity in the solution set maintaining the structural similarity with the parents. The results obtained by the proposed genetic algorithm substantiate the design of genetic operators for solving the QAP. Michalewicz et al. (1996) suggested the genetic algorithm-based solution method to implement the short-term unit commitment as a feasible alternative approach in a parallel machine environment. Rajasekhara et al. (1998) formulated a mathematical model for the FLP with the assumption that the rectilinear distance between pickup and drop-off points in a rectangular shape of the cell is restricted to the cell axes. Hamamoto and Salvendy (1999) proposed a simulation model for the pharmaceutical industry with the flexibility to define the multiple important criteria in case of a layout design. The experimental results obtained from the genetic algorithm- based solution approach outperform existing computer layout algorithms such as Craft, CORE LAP, and BLOC PLAN as well as human designers in maximizing the throughput rate and minimizing the traveling time/trip. Zhang et al. (2002) examined a multi-level layout of a warehouse having distinct storage areas to store the distinct items for inventory requirements. Extensive computational experiments are conducted using GA to minimize the transportation cost in a planning period. Dunker et al. (2005) presented a genetic algorithm- based model to solve the dynamic facility layout problem of unequal department size. Liuand Li (2006) ameliorated the parameters of genetic algorithms to determine the optimal locations of the departments in the supply chain-oriented dynamic facility layout problem. The numerical analysis substantiates the genetic algorithms on the practicability in resolving FLP. Kulkarni and Shankar (2007) applied a two- stage sequential approach using the GA to minimize the intercellular and the intracellular distance of material flow. The proposed methodology considerably simplifies the mathematical complexity of the QAP of manufacturing cells layout. Ripon et al. (2012) formulated a multi-objective mathematical model to reflect the real-world manufacturing scenario of the job shop scheduling problem and the facility layout problem. The experimental results approve the efficacy of the GA- based solution approach in the production of non-dominated best solutions. Pourvaziri and Naderi (2014) introduced a hybrid GA that generates the initial solutions for a Dynamic Facility Layout Problem (DFLP). The results compared with 11 other algorithms reported in the literature approve the efficiency of the proposed solution method. Paes et al. (2017) addressed the GA-based solution approach for the unequal area of FLP and aimed to locate rectangular facilities in unlimited floor space. The objective is to minimize the summation of distances weighted by the material flows among the facilities. The solution structure is

imposed to decompose the problem efficiently by limiting the facility to cross the X or Y-axis. Peng et al. (2018) suggested a robust layout for the assignment of transport devices in a dynamic facility layout using GA. Besbes et al. (2020) arranged the facilities in the boundaries considering the overlapping, aisle, and obstacle in planner sites. The objective is to minimize the material handling cost using an integrated GA. The literature cited above reveals several optimization algorithms for the formation of optimum facility layout design. The uses of the solution approaches are restricted to the formulation of the objective function. However, the major objective of minimum material flow in FLP is attempted by all the solution methodologies. The proposed research work aims at developing a GA-based solution approach and an embedded emendation heuristic to treat the problem of facility layout subjected to production- derived constraints. Next, the mathematical formulation of the FLP is also presented. A detailed solution methodology is elucidated and the computational practice of the results is also discussed.

III. RESULTS AND DISCUSSION

A. Mathematical Formulation The facility layout problem can be delineated as the determination of optimal locations for the facilities under the restrictions of shape and size considering the orientation for the pick-up/drop-off point for the material handling. The FLP in its simplest form can be formulated as a Quadratic Assignment Problem (QAP) having similar orientation and parameters for all the facilities considered in the equal-sized locations of a given plant layout. The Quadratic Assignment Model of the FLP addresses the optimal flow of material from one machine to another in an order to accomplish the machining operations by assigning each machine to one location in a given layout of the manufacturing system. The overall objective is to minimize the total material handling associated with per unit distance flow between facilities considering production volumes and production routing. The following notations are used for the development of the objective function: Mathematical Model

$$\text{MinTF}=\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{\substack{j=1\\ j\neq l}}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} F_{ij} * D_{jl} * X_{ij} * X_{kl} \quad \ldots(1)$$

$$\sum_{j=1}^{n} X_{ij} = 1 \;\; \forall i = 1, \ldots, n \quad \ldots(2)$$

$$\sum_{i=1}^{n} X_{ij} = 1 \;\; \forall j = 1, \ldots, n \quad \ldots(3)$$

$$X_{ij} \in \{0,1\} \quad \ldots(4)$$

The equation set: (1) Delineates the objective function that minimizes the summation of material flow multiple of the distance factor between facilities of a possible layout; (2) Confirms the assignment of each facility at one location; (3) Ensures that each location is equipped with one facility; and (4) Represents the logical binary non-negative restriction on decision variables.

B. Genetic Algorithm-Based Solution Schema The generation of a random set of initial solutions is the first operator of the GA. The chromosome representative of an individual solution is a single-level string describing the number of locations of the facilities. Table I illustrates the solution schema encoding. The objective function (Equation 1) calls the initial solution set as an input to compute the summation of material flow for each solution. The chromosome's fitness in the GA is mapped as a maximization of the objective function value. Thus, the necessary transformation of the objective function to the fitness for the minimization of the FLP is carried out in the following manner:

TABLE I. ENCODING OF CHROMOSOMES

Location	1	2	3	4	5	6	7	8	9	10	11	12
Facility	3	1	10	12	4	9	7	2	8	5	6	11
Parent	3	1	10	12	4	9	7	2	8	5	6	11

$$F_i = \frac{TF_{min}}{TF_i}, i = 1,2,\ldots.k \quad TF_i > 0 \qquad \ldots(5)$$

The equation set 5 describes the evaluation of fitness value Fi in the population set K equals the fraction of the smallest objective function value TFmin and the objective function TFi of the string i in the current generation. The relative performance of the fitness value is evolved by playing the binary tournament in the population set for the selection of elite parents to generate the offspring. Once the mating pool is formed, individual solutions are randomly paired and genetic operators are applied with a certain probability value of crossover and the mutation to generate the offspring. The crossover operator produces offspring by randomly exchanging the elements of the parent solutions. The crossover and mutation genetic operators may distort the results by violating the constraints of the assignment of each facility to one location in the given layout. The distorted offspring having a deficit of facility types are emendated to preserve the types of the facility as per constraints equation sets 2 and 3. The GA stops functioning after the evolution of the pre-specified number of generations

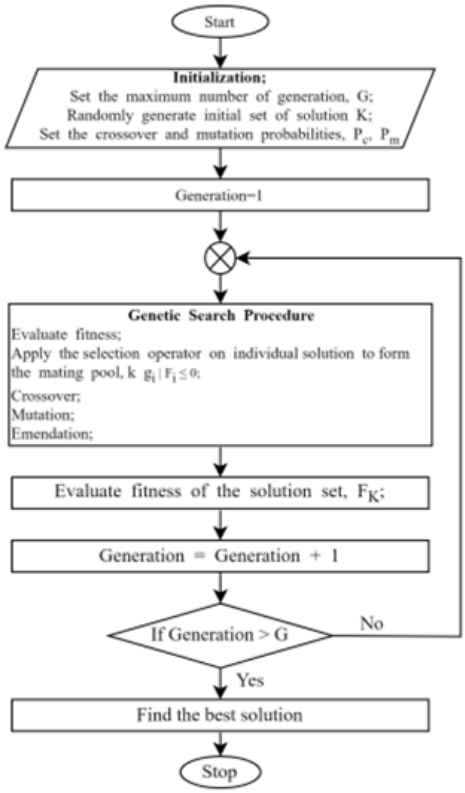


Fig. 1. Flow Chart of the Genetic Algorithm-Based Heuristic

IV. CONCLUSION

This paper reports the GA-based solution approach for the Quadratic Assignment Model of the FLP. A mathematical model has been developed to evaluate the pattern of material flow in a facility layout subjected to production-derived constraints. The computational analysis has been presented to evolve the optimal facility layout to minimize the weighted flow of material among the facilities. The efficiency of the proposed model has been examined on the benchmark problems reported in the literature. The comparative results exhibit the efficacy of the modest genetic operators for the facility layout design, which is an allusion to the probability of amelioration in the local search space of the genetic algorithm to achieve a near-global solution. The proposed model can also be extended to answer several complex problems with nonlinear constraints such as cost risk and trade-offs.

V. REFERENCES

[1]. Azadivar F and Wang J, "Facility Layout Optimization Using Simulationand Genetic Algorithms", International Journal of Production Research, Vol. 38, No. 17, pp. 4369-4383, 2000.

[2]. Banerjee P and Zhou Y, "Facilities Layout Design Optimization with SingleLoop Material Flow Path Configuration", International Journal of ProductionResearch, Vol. 33, No. 1, pp. 183-204, 1995.

[3]. Besbes M, Zolghadri M, Affonso R C et al., "A Methodology for SolvingFacility Layout Problem Considering Barriers: Genetic Algorithm Coupled withA* Search", Journal of Intelligent Manufacturing Vol. 31, pp. 615-640.4, 2020.

[4]. Chae J and Peters B A, "A Simulated Annealing Algorithm Based on aClosed Loop Layout for Facility Layout Design in Exible Manufacturing Systems",International Journal of Production Research, Vol. 44, pp. 2561-2572, 2006.

[5]. Drezner Z, "A Heuristic Procedure for the Layout of a Large Number ofFacilities", Management Science, Vol. 33, , pp. 907-915, 1987.

[6]. Dunker T, Radons G and Westkamper E, "Combining EvolutionaryComputation and Dynamic Programming for Solving a Dynamic Facility LayoutProblem", European Journal of Operational Research, Vol. 165, pp. 55-69, 2005.

[7]. Hamamoto S, Yih Y and Salvendy G, "Development and Validation ofGenetic Algorithm-Based Facility Layout—A Case Study in PharmaceuticalIndustry", International Journal of Production Research, Vol. 37, pp. 749-768, 1999.

[8]. Kazerooni M, Luonge L and Abhary K, "Cell Formation Using GeneticAlgorithms", International Journal of Flexible Automation and IntegratedManufacturing, Vol. 3, Nos. 3&4, pp. 283-299, 1995.

[9]. Kulkarni P C and Shankar K (2007), "Genetic Algorithm for Layout Problems inCellular Manufacturing Systems", Ieee International Conference on IndustrialEngineering and Engineering Management. [10] Liu X and Li X, "An Improved Genetic Algorithms-Based Approach onSupply Chain-Oriented Facility Layout Scheduling System", Proceedings of the6th World Congress on Intelligent Control and Automation, June 21-23, Dalian,China, 2006

[10]. Michalewicz Z, Genetic Algorithms C Data Structures Z Evolution Programs,Springer, Berlin, Germany, 1992.

[11]. Michalewicz Z, Dasgupta D, Le Richi R G and Schoenauer M, "EvolutionaryAlgorithms for Constrained Engineering Problems", Journal of Computers &Industrial Engineering, Vol. 30, No. 4, pp. 851-870, 1996.

[12]. Nugent C E, Vollman T E and Ruml J, "An Experimental Comparison ofTechniques for the Assignment of Facilities to Locations", Operations Research,Vol. 16, pp. 150-173, 1998. [14] Paes F G, Pessoa A A and Vidal T, "A Hybrid Genetic Algorithm withDecomposition Phases for the Unequal Area Facility Layout Problem", EuropeanJournal of Operational Research, Vol. 256, No. 3, pp. 742-756, 2017.

[13]. Peng Y, Zeng T, Fan L, Han Y and Xia B, "An Improved Genetic Algorithm-Based Robust Approach for Stochastic Dynamic Facility Layout Problem", DiscreteDynamics in Nature and Society, doi.org/10.1155/2018/1529058, 2018.

[14]. Pourvaziri H and Naderi B, "A Hybrid Multi-Population Genetic Algorithmfor the Dynamic Facility Layout Problem", Applied Soft Computing, Vol. 24, pp. 457-469, 2014.

[15].Rajasekharan M, Peters B A and Yang T, "A Genetic Algorithm for FacilityLayout Design in Flexible Manufacturing Systems", International Journal ofProduction Research, Vol. 36, No. 1, pp. 95- 110, 1998. [18] Ripon K S N, Glette K, Hovin M and Torresen J, "A Multi-ObjectiveEvolutionary Algorithm for Solving Integrated Scheduling and Layout PlanningProblems in Manufacturing Systems", IEEE Conference on Evolving and AdaptiveIntelligent Systems, 2012.

[16].Singh S P and Sharma R R K, "Two-Level Modified Simulated Annealing-Based Approach for Solving Facility Layout Problem", International Journal ofProduction Research, Vol. 46, No. 13, pp. 3563-3582, 2002.

[17].Suresh G and Sahu S, "A Genetic Algorithm for Facility Layout",International Journal of Production Research, Vol. 33, No. 12, pp. 3411-3423, 1995.

[18].Tavakkoli R and Shayan E, "Facilities Layout Design by GeneticAlgorithms", Proceeding of 22nd International Conference on Computer andIndustrial Engineering, pp. 415-418, 1997.

[19].Tompkins J A and White J A, Facilities Planning, 2nd Editioin, John Wiley,New York, 1996.

[20].Venugopal V and Narendran T A, "Genetic Algorithm Approach to the Machine-Component Grouping Problem with Multiple Objectives", Journal ofComputers and Industrial Engineering, Vol. 22, pp. 469-480, 1992. New Horizon College of Engineering, Department of Computer Science and Engineering 4

[21].Wang M J, Hu M H and Ku M H, "A Solution to the Unequal Area FacilitiesLayout Problem by Genetic Algorithm", Computers in Industry, Vol. 56, No. 2, pp. 207-220, 2002.

[22].Wu Y and Appleton E, "The Optimisation of Block Layout and AisleStructure by a Genetic Algorithm", Computers & Industrial Engineering, Vol. 41, No. 4, pp. 371-387, 2002.

[23].Zhang G Q, Xue J and Lai K K, "A Class of Genetic Algorithms for Multiple-Level Warehouse Layout Problems", International Journal of Production Research,Vol. 40, No. 3, pp. 731-744, 2002.

[24].Zhang Y, Zhu X and Luo Y, "Applying Genetic Algorithms to Task Planningof Multi-Agent Systems", Proceeding of 22nd International Conference onComputer And Industrial Engineering, pp. 411-414, 1997

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Scrutinization on Security and Infrastructure of Virtualization

Shrinidhi M.H., Sneha G.P., Prafulla, Sanjivani
Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

In existent days, the cloud computing grows in high charge with the assist of latest technological advancement. It affords the on call for and occasional price offerings to the purchaser. Cloud computing distributes the services to diverse region based on the demand. The services can go upto various layers and vicinity which leads to the security hassle. The existing security system presents reliable answers which are not sufficient to satisfy the reason. There are three forms of safety available within the cloud specific infrastructure protection and platform security. The proposed evaluation offers the whole information about these security and its problems. Information security control method has numerous steps for securing the data around the cloud internet-work. Virtualization level protection is analyzed with extraordinary parameter primarily based on the number of problems present in the cloud. Authentication kinds are also identified and analyzed over security level. Numerous styles of cloud framework and vulnerabilities are assessed as a way to gain higher protection.

Keywords: Cloud Computing, Data Security, Virtualization, Authentication, Cloud Services, Hypervisor.

I. INTRODUCTION

Despite the much advantages of virtualization, it also provides a facility to regulate and manage IT operations along with the cost reduction and improved efficiency. Security of information is the main concern in any IT environment. Current researches describe benefits of virtualization technology related to Information security. It is recommended that information safety procedures must be synchronized with business as virtualization deployed. Information Security (IS) requires proper audit and monitor system. The Information Technology Infrastructure Library (ITIL) is worldwide accepted standard in recent years related to information security and hence is greatly proper to serve as guide for the deployment and assessment of information security (IS) processes of virtualized environment. The primary aspect of virtualization and its management is the cloud and its functionalities. Cloud computing refers to the creation of a computer with a collection of visual computer resources. The cloud can handle a wide range of tasks, allow workloads to be depleted when required by the immediate provision of virtual or virtual machines, supports models of obsolete, recoverable, highly tolerant systems and allows workloads to recover from hardware or software failures and equal shares.

The idea is to move a desktop computer to a service-centered platform using clusters of servers and a large database in data centers. Different virtualization methods can be used for a variety of system layers including hardware, desktop, operating system, software, memory, storage, data and network. Complete virtualization is type of virtualization Hardware that involves complete extraction of beneath hardware and provides better performance by placing more work on each physical system. Full virtualization can be divided into two types: i) bare metal virtualization and ii) hosted virtualization. The bare metal method is widely used in server applications in large computer programs such as Cloud computing as it provides better more efficiency, durability and speed. metal based virtualization is commonly used in Cloud.

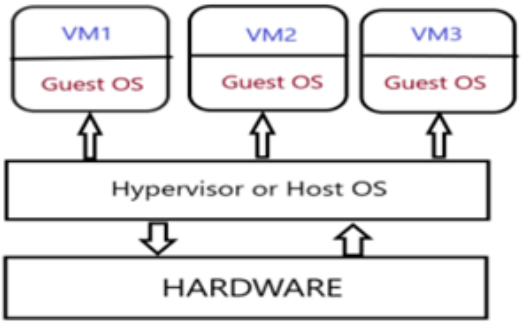


Figure 1: Layers of Virtualisation

The virtualization system consists of a host operating system, a hypervisor, and a guest operating system. Host is the basic hardware of a virtual reality system that provides computer resources (physical activity cards, CPUs, memory, control holes, operating system, third-party software, hypervisor). All of this allows the host to contain one or more Visible Machines (or partitions) and to share visual resources with them.

II. INFRASTRUCTURE

In most cases, data centres contain a large number of web servers, the majority of them are inactive since the job is limited to only a few web servers on the network. This results in the waste of costly resources, equipment, electricity, maintenance, and other components of the business. By segmenting a physical server into online servers, server virtualization aids in increasing resource utilisation. These virtual web servers have the appearance and behaviour of a private physical web server with its own operating system and applications. Virtualization is a concept that is used in almost every IT architecture to assist boost the capability of physical equipment. It aids in maximising the use of current resources. As a result, the company's total costs are reduced. Virtualization technologies and operating systems are used by businesses to create virtual environments. Machines, networks, desktops, and web servers are all examples of computer hardware. Virtualization architecture simplifies IT, allowing you to operate more efficiently and cost-effectively. It delivers a slew of benefits, including cheaper hardware costs, improved disaster recovery solutions, more IT agility, improved performance, and a rapid source timetable. It's critical to comprehend the various forms of server virtualization before implementing it. Based on their technological requirements, businesses and services might choose the type of virtualization solution they want. The following are some of the other advantages of virtual infrastructure: Downtime on the fringes Virtualized computing resources are simple to specify and install. Downtime due to risk management becomes considerably more fluid and adaptive as a result of this. Instead of replacing, repairing, or cleaning up a damaged or virus-infected computer, you may instantly build a new system and restore its contents from your backup service.

The reduction in operating expenses is one of the most well-known and recognised benefits of virtualization. The capacity to share resources reduces the amount of physical gear that a corporation has to acquire. Having a single web server that can arrange several servers, or a single desktop that can run a variety of various systems, means less money spent up front. The cost reductions, though, aren't confined to hardware. With fewer equipment on your network, you'll use less energy and spend less time maintaining it. For your firm, this could only imply one thing: cost reductions!

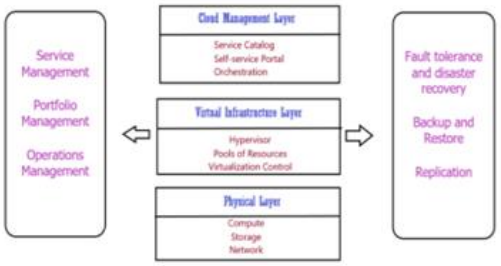


Figure 2: Infrastructure of Virtualisation

III. SECURITY

Cyber assaults are an increasing day by day. When attacks occur, it is important to get systems back up. Virtualization helps resiliency with the aid of using supporting to shield the OS kernel, application, and secrets and techniques from malware. Virtualization additionally permits the introduction of more than one remoted paintings environments, which might be used, for example, to permit separation among paintings and private hobby for higher protection, records privacy, and compliance. Virtualization protection is the collective measures, techniques and methods that make certain the safety of a virtualization infrastructure addresses the safety problems confronted with the aid of using the additives of a virtualization surroundings and techniques via which it may be prevented. Typically, virtualization protection may also consist of methods such as:

- Implementation of protection controls and techniques granularly at every digital machine.
- Securing digital machines, digital community and different digital equipment with assaults and vulnerabilities surfaced from the underlying bodily device.
- Ensuring manage and authority over every digital machine. Virtual security, or security virtualization, refers to softwarebased security solutions designed to work within a virtual IT environment. This differs from hardware-based network security, which stands and works on devices such as firewall, routers, and switches. New Horizon College of Engineering, Department of Computer Science and Engineering 3 In contrast to hardware-based security, virtualized security is flexible and dynamic. Instead of being attached to a device, it can be installed anywhere in the network and is usually cloudbased. This is the key to virtualized networks, where operators exchange functions and applications dynamically; visual security allows security services and operations to move with those loads created by power. Cloud security considerations are also important for virtualized protection. The flexibility of virtual security helps to protect mixed and cloudy environments, as data and workloads move to a more complex ecosystem that includes multiple vendors

Best practice for a secured virtualization environment:

1. Network Security

- Eliminate loopholes into the machine with the aid of using disconnecting any inactive NIC.
- Secure the host platform that connects visitors and hypervisors to a physical network with the

aid of using setting up logging and time synchronization, putting matters in vicinity to alter customers and groups, and placing record permissions.
- Use authentication and encryption on every packet to stable IP communications among hosts.

2. Disaster Recovery
- Have a right alternate manage in order that the principle web page and the backup web sites are saved as identical as possible
- PEN check and auditing must be one after the other executed in your DR web page and the principle web page however with the identical frequency and significance.
- Logging and different statistics sourced from the DR web page must be handled with the identical significance as those from your primary site.

3. Separation of responsibilities and Administrator access
- Server directors must be provided, specifically, with credentials of the respective servers they're in charge of.
- Admins must take delivery of the strength to create new VMs however now no longer to adjust already present VMS.
- Every guest OS must be assigned a completely unique authentication except there may be a compelling purpose for 2 or extra guest OS to apply the identical credentials.
- Contrary to not casual thought, safety employees have determined out that the bigger the virtualized environment, the less complicated it allocates duties throughout functions. An admin can't perform the complete control manner singlehandedly.

4. Virtual Machine Security
- Management networks related to hypervisors should not be be used to store VMs
- Using processor-in depth screensavers on physical servers crush the processor had to serve the VMs
- Only create VMs as in step with the requirement. Unused VMs within side the surroundings can shape capability access factors for black hats.
- The kennel or host resources, inclusive of garage networks, need to be without problems accessed through VMs.
- Disable all unused ports, inclusive of USB ports present on VMs.
- Encrypt records being conveyed among the Host and VM.

5. Backups
- No root accounts ought to be used for backups.
- Disk backups are similarly as essential within side the virtualized surroundings as they may be within side the conventional one.
- Perform a complete gadget lower back as soon as per week and common or each day backup of OS and information
- Encrypt each information dispatched to a disaster restoration over the network

IV. IMPLEMENTATION OF SECURITY

There are many ways in which security is implemented with various features and types of virtual security, including network security, application security, and cloud security. Other virtual security technologies are truly updated, virtual versions of traditional security technologies. Another new technology built into it and a virtualized network fabric.

Other common types of virtual security features include:

- Segmentation, or certain services available only for specific applications and users. This usually takes the form of traffic control between different network segments or categories.
- Micro segmentation, or applying certain security policies at the workload level to create secure granular areas and reduce the attacker's ability to navigate the network. Microsegmentation divides the data center into segments and allows IT teams to define security components for each segment, which strengthens the data center's resistance to attack.
- Isolation, or separating individual workloads and applications from the same network. This is very important in a multi-community cloud environment, and can be used to separate virtual networks from virtual infrastructure, to protect infrastructure from attack.

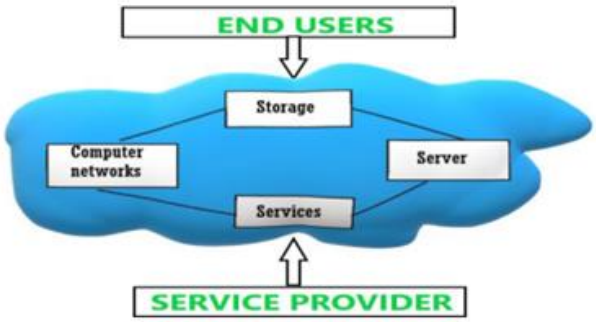


Figure 3: Security of Virtualisation

Working of virtual security:

Virtual security can take over the functions of common security software and use them with software. In addition, virtualized security can also perform additional security functions. These activities are possible because of the benefits of virtualization, and are designed to address specific security needs of the virtual environment. For example, a company may incorporate security controls (such as encryption) between the application layer and the underlying infrastructure, or use strategies such as subdivision to minimize potential attacks.

Virtual security protection can be used as an application directly to the bare metal hypervisor (a position we can upgrade to provide active application monitoring) or as a hosted operating system. In any case, it can be used immediately where it is most effective, as opposed to physical security, which is attached to a particular device. Normal physical protection is based on hardware, and as a result, it is unchanged and static. The standard method depends on the devices used in the strategic areas across the network and usually focuses on protecting the network cycle (as a standard protection wall).

However, the perimeter of the cloud-based virtual reality network has porous and loads of operations and applications are created by force, which increases the potential for attack. Traditional security is also very much dependent on protocol filtering and port, a method that does not work in a virtual environment where addresses and ports are provided dynamically. In such a case, protection based on traditional hardware is not sufficient; a cloud-based network requires virtual security that can move between the network and loads of functions and applications.

V. RESULTS AND DISCUSSION

Modern organizations need to protect their physical environment against a variety of threats. Key strategies include keeping all software up to date, using AV software, following the best configuration procedures and performing regular user training. But even with the best protection, some threats will pass, so it is important to invest in security tools that can track changes and logos to help you maintain security at all levels, at all times.

Infrastructure components	Functional Description
Virtualized computer:	This component offers the same capabilities as physical servers, but with the ability to be more efficient. Through virtualization, many operating systems and applications can run on a single physical server, whereas in traditional infrastructure servers were often underutilized.
Virtualized storage:	This component frees organizations from the constraints and limitations of hardware by combining pools of physical storage capacity into a single, more manageable repository.
Virtualized networking and security:	This component decouples networking services from the underlying hardware and allows users to access network resources from a centralized management system.
Management solution:	This component provides a user-friendly console for configuring, managing and provisioning virtualized IT infrastructure, as well automating processes.
Host:	A virtualization layer that manages resources and other services for virtual machines. Virtual machines run on these individual hosts, which continuously perform monitoring and management activities in the background.
Hypervisor:	A software layer that enables one host computer to simultaneously support multiple virtual operating systems, also known as virtual machines.
Virtual machine:	These software-defined computers encompass operating systems, software programs and documents. Managed by a virtual infrastructure, each virtual machine has its own operating system called a guest operating system.

Virtual security is now effectively required to comply with the complex security requirements of a virtual network, and is much easier to adapt and work than with standard physical security. Here are some of its specific benefits:

- Cost effective: Virtual security allows the company to maintain a secure network without significant cost overruns on expensive proprietary computer hardware. The price of cloud-based security services is often determined by usage, which could mean more savings for organizations that use resources more efficiently.
- Adaptability: Virtualized security functions can track workloads anywhere, which is important in a virtual environment. It provides protection across all data centers and multi-cloud and hybrid computing environments, allowing the organization to take full advantage of virtual reality while keeping data secure.
- Operational efficiency: Faster and easier to use than hardware-based security, virtual security does not require IT teams to set up and configure multiple hardware components. Instead, they can set up security systems using central software, allowing faster scaling. Utilizing software to advance security technologies and allows security operations to be automated, which frees up more time for IT teams.
- Compliance: Indigenous hardware-based security is static and unable to comply with the requirements of a virtual network, making visual security a requirement for organizations that need to maintain compliance.

Operational efficiency:	Quicker and easier to deploy than hardware-based security, virtualized security doesn't require IT teams to set up and configure multiple hardware appliances. Instead, they can set up security systems through centralized software, enabling rapid scaling. Using software to run security technology also allows security tasks to be automated, freeing up additional time for IT teams.

Security Control Roles	Functional Description
Network Security:	It ensures that proper security controls are in place for the virtualized network, by providing network zoning segmentation and policy enforcement.
Platform Security:	It supports the hardening of the overall platform, by defining configuration control points driven from security requirements, which includes security posture hardening of the various virtualization management layers.
Data Protection:	It provides security mechanisms for isolation and encryption in alignment with overall data protection requirements.
Access Control:	It provides policy decision and enforcement to ensure that authentication, authorization and accounting is implemented across the virtualization infrastructure for systems administrative access, aligned with IAM framework.
Security Management:	It provides security Management controls to account for detective, preventative, and reactive measures required for the virtualization infrastructure (e.g. logging/SIEM, vulnerability management/scanning intrusion detection)
Software Requirement:	Security virtualization uses software installed on virtual security networks to monitor workloads, applications, and access to VMs. The security can also manage security policies for access to virtual networks and workloads themselves.

VI. CONCLUSION

Various security methods have been presented in the literature to address assaults on virtualization environments. The security solutions for each component of the virtualization architecture are discussed in this section. The threats outlined in section 3 can be neutralized or at least the impact of such attacks on the virtualized environment may be limited by applying these security measures Cloud security is impossible to maintain unless the virtualization environment is safe. Although other virtualization technologies exist, bare metal virtualization is typically utilised for server virtualization in big computing systems such as Cloud. The overall architecture of bare metal virtualization is presented in this article, as well as the security concerns of its many components. Different attacks on the service provider, hypervisor, virtual machines, guest operating system, and disc images might damage the cloud virtualization environment. In the study, attack possibilities against these components are addressed. General needs for virtualization security and other existing security schemes that give security to virtualization environments have also been examined in order to provide security to the virtualization environment.

As a result, a systematic analysis of virtualization security in the Cloud provides a comprehensive picture in which security needs, threats, and solutions are correlated. Addressing these security concerns will lead to more in-depth study into securing the Cloud virtualization environment. In the future, an assessment criterion must be established by which we can evaluate the efficacy of virtualized security solutions against particular assaults, security needs, attacks, and solutions are all mapped out in an organised analysis. Addressing these security concerns will lead to more in-depth study into securing the Cloud virtualization environment. In the future, an assessment criterion must be established by which we can evaluate the efficacy of virtualized security.

VII. REFERENCES

[1]. Mathew Ponny, “Virtualization Essentials”, 2nd ed.

[2]. N L Udaya Kumar, Siddaramaiah S R. 2015. International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCACCT-2015 Conference Proceedings.

[3]. Prashanta Kumar Das (Government Industrial Training Institute Dhansiri, India) and Ganesh Chandra Deka (Government of India, India), in Computer Science & IT for 2019.

[4]. Rajkumar Buyya ,JamesBroberg, Andrzej M. Goscinski, "Cloud Computing: Principles and Paradigms" 5th ed.

[5]. Rajkumar Buyya, Christian Vecchiola,,S. ThamaraiSelvi ,"Mastering Cloud Computing: Foundations and Applications Programming" 4th ed.

[6]. Edouard Bugnion, Hardware and Software Support for Virtualization (ebook)

[7]. Virtualization Security, GRD JOURNALS,Rishabmishra, Abhinav Mishra

[8]. A survey of virtualization security, International Journal of Scientific & Engineering Research,SomayehSobatimoghadam

[9]. Virtual Infrastructure (Journey from Physical Servers to the Cloud), International Journal of Computer Trends and Technology, Subash Thota

[10]. Virtualization In Cloud Computing,Ramandeepkaur,Sumithchopra, International Journal of Scientific Research in Computer Science, Engineering and Information Technology

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

The Entirety of The Voice Assistant

Nishitha Tanukunuri, Atharva K, Saloni B, Vijay Kumar Reddy P
Department of Computer Science and Engineering, New Horizon College of Engineering Bengaluru, Karnataka, India

ABSTRACT

Digital voice assistants such as Siri, Alexa, etc are complex and futuristic artificial intelligence (AI) technology. These Voice assistants can be used to do basic personal task management duties as well as more advanced capabilities and connected-device integrations. However, the functional and topical use of a digital voice assistant varies from person to person. As a result, the respondents' contextual knowledge is reflected in this study. Only a little amount of empirical evidence of client satisfaction with digital voice assistants exists at the moment. To investigate this research gap, PLS-SEM was largely employed to analyze 245 survey responses. While Siri users made up the majority of the sample (72 percent), additional digital voice assistants (28 percent) were also identified and included. The findings also revealed that customer expectations and confirmation of those expectations have a very favourable and considerable impact on the customer satisfaction when it comes to digital assistants. From a practical standpoint, this research shows that customers' expectations are being precisely met through the digital voice assistant contact experience. As many of the multinational corporations continue to integrate digital voice assistants into many or all of their operations, they must ensure that consumers are effectively assisted and that customers understand what to expect from the firm's interactive experience. Regardless of the circumstances, consumers have always prioritized the validation of these expectations.

Keywords—Digital voice assistant, customer satisfaction, expectations confirmation theory, privacy concerns, artificial intelligence, progressive capabilities.

I. INTRODUCTION

Consumers and businesses across the globe are quickly adopting digital smart speakers like Google Assistant, Siri, and Alexa, along with many others. Cohesive artificial intelligence technologies with speech capabilities are known as virtual assistants. Also commonly referenced as conversation-enabled applications. They are thought of as dynamic systems with superior natural language understanding and the ability to learn customer preferences [6]. These virtual systems deliver assistance to users by addressing to questions in the user's natural language, offering recommendations and implementing actions leveraging inputs such as a person's voice, vision (videos, photographs, etc.), and other circumstantial information [2] . According to Statista Research, there are currently 4.25 billion digital voice assistants in use throughout the world. This growth should be reflecting that by the year 2023,

- More than 50% of all searches done online will be voice-activated.
- 55% of all households in the US are expected to own a smart speaker in 2022.
- More than 20% of all mobile queries are currently done with speech search.
- Nearly 45 million Americans currently own a smart speaker.
- Up to 58% of people have used voice search to find information about a local business.
- 72% of the people who use voice search devices claim that virtual assistants have become part of their daily routines.
- More than 43% of smart speaker owners use the technology to shop.
- 32% of all marketing organizations are using smart speakerslike Alexa and Siri. Multinational corporations are implementing integration initiatives that include conversational or chatbot consumer interfaces on websites, social media, mobile applications, inventory management, and banking automation, among other things. This piece of technology is also being integrated into collaborative work activities, such as exchanges amongst conversational agents and individuals for call - center support. Given the significant investment that various corporations are making in virtual voice assistant technology, as well as the re-design of fundamental production and customer service processes to accommodate this technology, proof that users trust and appear to be content with it is required. The following data has been sourced from statistica research [10].

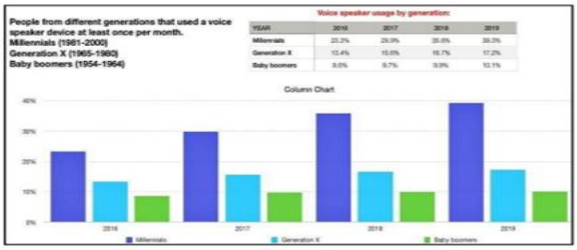


Fig.1. Voice Search Demographic Statistics[10]

INTERPRETATION: From the above figure, we can conclude that there is an increasing amount of usage of voice assistants in millenials and the least being baby boomers.We can see that as the years go by there has been a slow increase of usage in generation X.

Data-driven discoveries incorporating previously undiscovered patterns, correlations, and supplementary illuminating personal interactive insights are now attainable thanks to recent breakthroughs in machine learning and deep skill set linked with digital voice assistants [9].

II. TECHNOLOGY

A. Working Of Google Assistant

Google Assistant instantly processes the words that you speak and understands exactly what you are trying to convey. It then responds by constructing clear sentences just like a human being would, based on what it has learnt. Google Assistant processes all the inputs and comes up with an output. Google Assistant also benefits from the machine learning technology, which primarily focusses on systems that teach machines to learn from experience and make the necessary decisions accordingly, rather than being programmed for every single task [7].

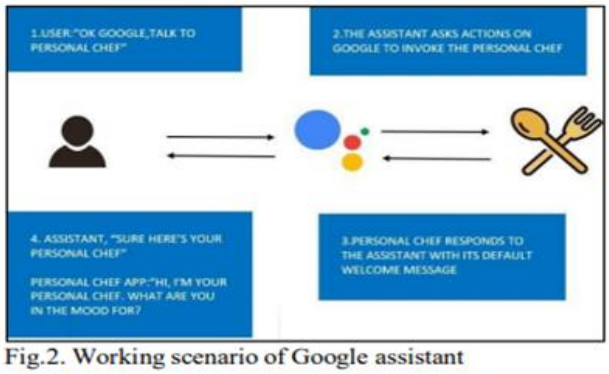


Fig.2. Working scenario of Google assistant

INTEPRETATION: The above figure depicts the working of google assistant. It shows how the user interacts with the voice assistant to get information like talking to the personal chef.

B. Working Of Apple Siri

Once the Siri microphone button has been touched, whatever is said isrecorded, is compressed and then sent to Apple's data centres where Apple hosts Siri's Nuance speech to- text and AI-like natural language processing engine. Siri then figures out what has been said by the user. Depending upon the request sent by the user, Siri creates a particular voice response and either sends it back to the iPhone or executes queries and sends the voice response file and data back to the handset. The iPhone's Siri Digital Assistant is provided "life" by vocalizing answers and, if merited, presenting information obtained via the Siri Backend services[1] (e.g., Google search results) or via the customer's iPhone resources (e.g., Map app).

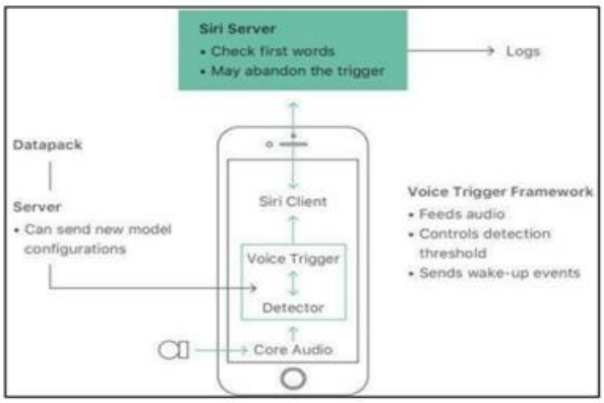


Fig.3. Working scenario of Siri

INTERPRETATION: The above figure depicts the working of apple Siri from the client which has voice trigger framework and gives request to the Siri server.

C. Working Of Amazon Alexa

Amazon Alexa works by processing requests or commands like through a machine learning technique called Natural Language Processing, NLP for short. In simpler words, Natural Language Processing is simply the manipulation of forms of natural language like text, audio, speech using algorithms. It all begins with the concept of Signal Processing, put simply, Signal Processing is simply the technique that helps speech-based systems to get rid of noise which prevents accurate processing of data [8].

After successfully processing the incoming signals, Alexa performs a wake word detection, during this process Amazon Alexa searches the target signal for an activation word, usually Alexa, Hey Alexa!Alexa's Neural Networks are trained by using supervised machine learning algorithms.

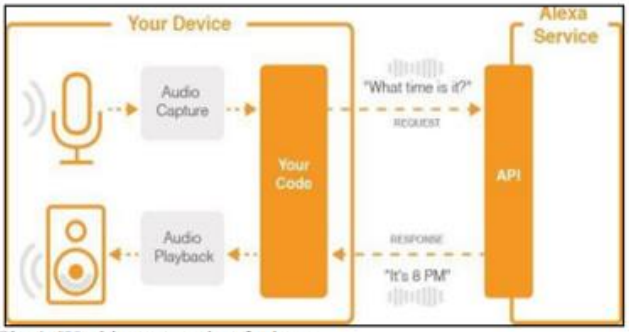


Fig.4. Working scenario of Alexa

INTERPRETATION: The above figure depicts the working of Amazon Alexa which uses a set of APIs for a request – response action.

III. ARCHITECTURE

A. *Architecture Of Google Assistant*

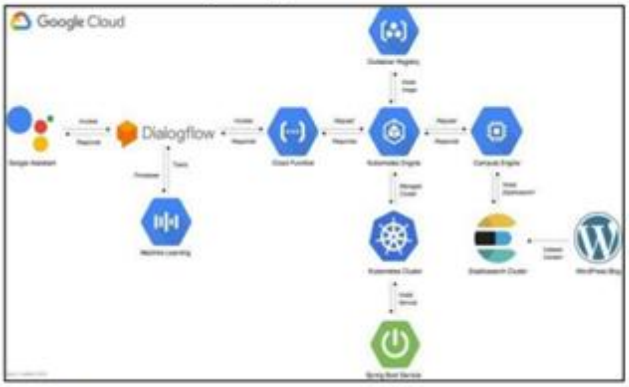


Fig.5. Descriptive architecture of Google assistant

INTEPRETATION: The above figure depicts the architecture of google assistant which includes machine learning concepts etc.

Google Assistant Service exposes a low-level API that lets you to directly manipulate the audio bytes of an

Assistant request and response. Bindings for this API could be generated for languages like Node.js, Go, C++, Java for all platforms that support gRPC. Reference code is provided in Python for audio capture, audio playback, and conversation state management.

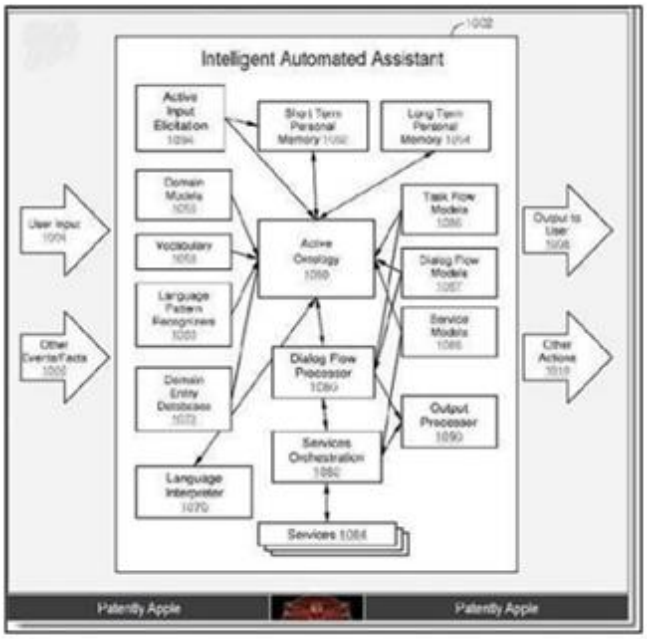


Fig.6. Descriptive architecture of Apple Siri

Interpretation: The above figure shows the descriptive architecture of Apple Siri which uses user input and other events/facts to give output to user or other actions.

B. Architecture Of Apple Siri

The Apple Siri works by using technologies such as Natural Language Processing, Speech recognition, Parsing, Lexical semantics, Word sense disambiguation and grammar [3]. It is of utmost importance to understand that Siri, unlike the Google's AI-driven assistant, performs tasks on the server-side, and not the client-side. Google's AI helper performs the initial voice recognition processes on the device itself before pulling the information from the server, Siri performs nearly all of its tasks on the server.

C. Architecture of Amazon Alexa

Fig.7. Descriptive architecture of Amazon Alexa

INTERPRETATION: The above figure depicts the detailed architecture of Amazon Alexa where user interacts with the voice assistant device which uses Amazon Voice Services.

Amazon Alexa enabled devices sends the user's instruction to a cloud-based service called the Alexa Voice Service [8].Think the Alexa Voice Service as the brain of all the Alexa enabled devices and perform all the complex operations such as Automatic Speech Recognition and Natural Language Understanding .

IV. CONCLUSION

All three virtual assistants described above are excellent at handling basic requests. Each of them can give you weather updates, point you in the direction of a decent restaurant, and seek up a variety of information on the internet. Alexa, on the other hand, is better for your home and office, while Google and Siri are better for when you're out and about. Siri is limited to Apple products, but Google has access to a much broader range of smartphones and home gadgets. In the world of voice assistants, there is still no one-size-fits-all winner [4] . Your best option is the one that excels in the categories that matter most to you and is compatible with the devices you want in the environment you want. And, given the quick speed of improvement in all of these technologies,

time spent with any of these voice assistants is difficult to go wrong [5] .

	AMAZON ALEXA	GOOGLE ASSISTANT	APPLE SIRI
GENERAL KNOWLEDGE	1st ☑	2nd X	3rd X
MUSIC & PODCASTS	2nd X	1st ☑	1st ☑
ENTERTAINMENT	1st ☑	2nd X	3rd X
ORDERING FOOD	2nd X	1st ☑	2nd ☑
ONLINE SHOPPING	1st ☑	1st ☑	2nd X
COMMUNICATIONS	2nd X	2nd X	1st ☑
DIRECTIONS	3rd X	1st X	2nd X
SMART HOMES	1st ☑	1st ☑	2nd X
AVAILABILITY	3rd X	2nd X	1st ☑
VOICE RECOGNITION	2nd X	1st ☑	3rd X
EXTENDABILITY	1st ☑	2nd X	2nd X
TRANSLATIONS	2nd X	1st ☑	2nd X
SCHEDULE	1st ☑	2nd X	2nd X
TOTAL NUMBER OF FIRST-PLACE WINS	6	6	4

Table-1: An understanding of the differences between the Virtual Assistants

INTEPRETATION: The above table shows a comparison between the three voice assistants in different aspects. The table concludes that Amazon Alexa and Google Assistant have the majority of first place.

Each of these virtual assistants has its own set of benefits and drawbacks. Given the current state of digital assistant acceptance and use, there is a scarcity of empirical research directly relevant to the consumer experience.

V. REFERENCES

[1]. http://www.jeffwofford.com/?p=817 (Must Read: Concise Siri summary)

[2]. Alpaydin, E. (2014). Introduction to Machine Learning.Cambridge, Massachusetts: MIT Press

[3]. https://machinelearning.apple.com/research?page=4&domain=Speech+and+Natural+Language+Processing

[4]. Alexa vs. Google Assistant vs. Siri: Which smart assistant is best | Tom's Guide (tomsguide.com)

[5]. How Amazon Alexa works? Your guide to Natural Language Processing (AI) | by Alexandre Gonfalonieri | Towards Data Science

[6]. Moar, J. (2019, February). The Digital Assistants of Tomorrow. Retrieved from https://www.juniperresearch.com/documentlibrary/white-papers/the-digital assistants-of-tomorrow

[7]. SCIplanet- The Technologies behind Google Assistant (bibalex.org)

[8]. Hyunji Chung, Jungheum Park, Sangjin Lee (2017). Digital Forensic approaches for Amazon Alexa ecosystem

[9]. Tom Brill, Richard Miller, Laura Munoz (2019). Siri, Alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications

[10].https://www.statista.com/statistics/973815/worldwidedigitalvoiceassistantinuse/#:~:text=In%202020%2C%20there%20will%20be,higher%20than%20the%20world's%20population

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Survey on Tracing IP Address Behind VPN/Proxy Server

Manikandakumar M, Nuthan KV
Department of Computer Science, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Criminals are not only employing cyber root to perform crimes, but they are also honing their skills with each passing transaction. The internet's anonymity encourages cyber thieves to carry out these attacks without fear. In most cases, cyber thieves will use proxy services to mask their genuine IP address and physical location when conducting fraudulent transactions. This is done in order to avoid being monitored by law enforcement and convicted. This study details the investigation of tracing IP addresses behind VPN / Proxy servers, as well as efforts to incorporate such technology into a commercial solution with the sole goal of preventing the majority of fraudulent transaction attempts. This method determines which IP address is behind the VPN/Proxy server and may be easily tracked using anonymous tools.

Keywords: Virtual Private Network, Proxy, Honeypot, Canary Tokens, Internet Protocol

I. INTRODUCTION

Virtual Private Network (VPN) is a term that refers to a private network that runs over a public network. It allows users to send and receive data from public networks that are connected directly to a private network. A proxy server is a server application that works as a middleman between a client who requests a resource and the server who provides that resource. Previously, this VPN/Proxy server could not be traced by third parties, however it may now be traced using anonymous methods. Many of the data can be obtained by using techniques such as the Man in the Middle attack and switching the data. We need a link between two locations and a static IP address for these attacks.

Anyone can obtain any information through the use of terminals such as Tor. Tor is a hybrid network that may be constructed in C or Python. In this universe, the Mac address is the most distinctive identity, and we may easily spoof it. So, in this article, we will look at how IP addresses may be traced behind VPN/Proxy servers using various tools, and this paper will have multiple sections that will provide information on this.

This paper is divided into five sections. The first section introduces VPN and explains how it works. Section II summarises existing research on IP addresses behind VPNs and how they are detected by various researchers. Section III delves into the research technique employed in this study, as well as the instruments employed and the recommended layout. The results are detailed in Section IV, which includes a comparison of the two honeypots. In section V, the conclusion and future work are presented.

II. EXISTING WORK

Over any public network, a Virtual Private Network (VPN) delivers private networks of resources and information. It allows a remote machine on network X to tunnel traffic to a gateway machine on network that would otherwise be inaccessible through the Internet. With an overall accuracy of 97.71 percent, OpenVPN will correctly identify the VPN. The VPN technology was chosen to focus on Neural Network, which is capable of handling VPN traffic and is utilized to provide encryption over a tunnel. As a result, it might be used to model a real-world dataset. The OpenVPN VPN technology was chosen as the subject of the studies, and it was discovered that the Neural Network was capable of identifying network traffic as either VPN or non-VPN traffic [1].

The network used unique identifying methods and had a secure and encrypted connection over a less secure network. The dedicated computer is a computer that sits between two endpoints and was created to give distributed systems structure and encapsulation [2]. It can make it easier for servers and web proxies to access WWW content and circumvent IP address blocking. The detection of malicious individuals involves using circuit cased networks such as Tor and SOCKS proxy to detect anonymous connections. It will achieve the conversion of network packets to protocol handshakes, as well as the detection of anonymous SSH and HTTPS connections [3].

The browser driver provides realism and support for web technologies in online tracking, while the browser manager ensures stability. Even third-party trackers employ two methods to classify advertising and non-advertise trackers: Easy List and Easy Privacy. To begin with, it will increasingly employ machine learning to automatically detect and classify trackers, as well as improve browser privacy capabilities [4].

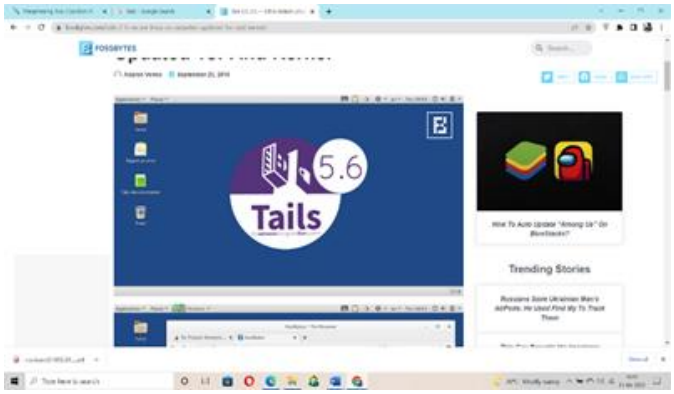


Figure. Tails OS

Tails comes with a unique set of software for encrypting files and internet transmissions, cryptographic signing and hashing, and other security-related tasks. It's already set up to use Tor and has a variety of connection choices. It aims to force all connections to use Tor and inhibits any attempts to connect outside of Tor. It is most used to protect the whole network like brute force attack.

Anonymous proxies are intermediary relays between the source IP address and the destination IP address, and they can play a significant part in cybercrime. Proxies for PHP are the most widely used. It can blog the servers with other programmers and detect the proxies that support the IP for more. Anonymous Proxy Servers The most significant risk is PHP obfuscation, which can be identified using the Base64 encoding method for security purposes. SSL CGI proxy is a downfall for packets in the network. As a result, it detects whether or not a proxy is in use [5].

Exploring the proxy method is a detection strategy that can lead to the creation of a multitiered proxy detection module in a commercial setting. This experiment resulted in a 97 percent detection rate and a minimal integration cost to prevent fraudulent transactions from non-VPN proxy connections [6].

The proxy server can be used by users' local computers, and it has the advantage of allowing

administrators to monitor and oversee internet usage from a single workstation, as well as protecting users from malware by limiting their access to only affected websites [7]. In order to avoid attacks, IP spoofing attackers can alter their IP address to a new IP address. It is utilized to maintain a trustworthy connection by suspending services. However, the method of prevention is dependent on the analysis of IP spoofing attacks [8].

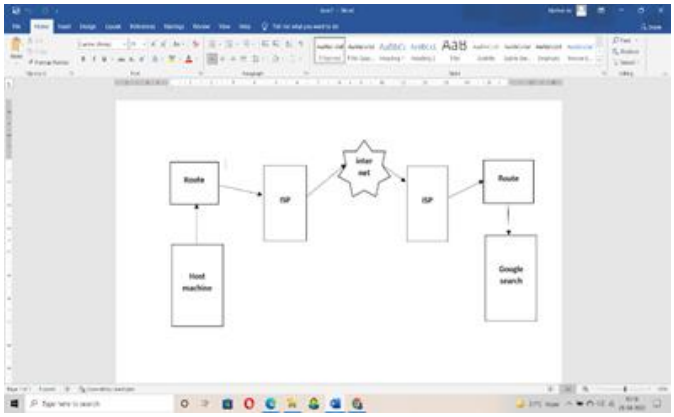


Figure. Routers & Switches are Connecting in IP

The number of IP addresses is used to analyse the stability and uniqueness of various endpoints that a user can connect to the internet for long-term tracking of the IP address and prospective insights into their lives[9]. Despite the fact that cookies are synchronizing, IP tracing remains a big danger to internet privacy for large-scale sites.

The system and web interface have been installed, and plans are in the works to integrate the Net craft's to determine which servers are providing fraudulent content and which are employing IP geolocation servers. It primarily employs routing information to obtain a large number of data sets in order to execute IP. The location was chosen as the target for tracing IP traceroutes [10].

Client IP addresses are hidden behind anonymity services, although they appear to be conventional residential IP addresses. It has identified commercial RESIP proxy service providers and investigated dubious host recruitment tactics. It is capable of detecting malicious RESIP proxy services that are running on a compromised host [11]. Even the detection was able to locate the additional malicious proxy connections with reasonable precision and accuracy.

In this IP spoofing attack, the reliance on internet service providers for IP traceback is reduced. It employs a proof technique that use a relative IP back-trace mechanism to determine the tracer's location [12]. The T classification of IP addresses can help with speed, simplicity of use, and robustness. Many apps employ IP geolocation to determine the Internet host, and they use four methods: data collecting, data cleaning, constraint calculation, and position estimation [13]. It is able to dissociate the geolocation findings from the experiment points and the density of landmarks. The challenge with safeguarding the anonymity of ISP networks in tracebacks is that there are more traffic routing paths to follow, which makes it easier to spot an attack.

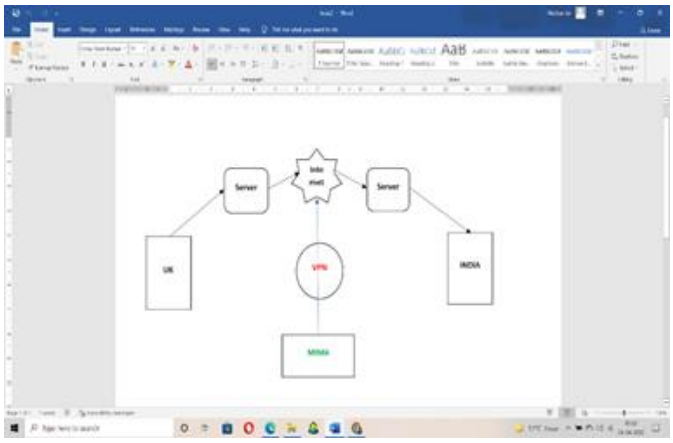


Figure. One location server connect to another location

In logging-based reconstruction, relevant packets information is saved in the reconstruction path. The algorithms for identifying and marking each length of the ISP server are built using simulation-based analyses, and the accuracy bits are used to detect and mark each length of the ISP server [14].

The browser driver provides realism and support for web technologies in online tracking, while the browser manager ensures stability. Even third-party trackers employ two methods to classify advertising and non-advertise trackers: Easy List and Easy Privacy. To begin with, it will increasingly employ machine learning to automatically detect and classify trackers, as well as improve browser privacy capabilities [15].

Dos attacks can make a service unavailable, and they can use a variety of methods to compromise a system in different locations. The attackers can execute coordinated attacks on the victims' computers using these compromised systems, which are referred to as zombies. It can prevent such protocols from exploiting network protocol flaws. The flaws took advantage of naive hosts to generate results that were unfit for routine use[16-18].

III. RESEARCH METHODOLOGY

A distributed measuring tool to monitor the free proxy ecosystem, as well as a methodology for recognizing proxies that actively manipulated traffic rather than passively repeating it. Unlike previous web proxy research, this article explores RESIP services and focuses on residential hosts as a means of traffic forwarding. In the beginning, legitimate users carried out all of their tasks using a standard web browser. Later, they switched to a low-latency anonymous browser, such as Tor. Because traffic patterns transit in an encrypted manner, identifying them in the Tor browser is extremely difficult. With the help of the Tor network, its identity will be masked. However, many attackers have recently begun to use Tor browsers. With a secure Tor network, botnet detection becomes more difficult. Because the Tor network is extremely secure, we do not advocate it for practical use. PTN (Private Tor Networks) and a trusted intermediary node are also suggested.

A regular browser displays the IP address that we can obtain. When I check up an IP address in the Tor browser, it displays the proxy address, which can be found in the leak DNS. To find the IP address, the attacker used SQL injection and dangerous attacks like brute forcing. Every organization's protection from any type of cyber assault has become a vital aspect of network security. To do so, we must first comprehend the traffic patterns that move over the network. This can be accomplished by deploying a large number of network intrusion detection systems. Honeypot is one of them. Honeypot aids in communicating with attackers as well as acquiring various information about network attacks. Because honeypots send data in small amounts, security experts can decipher it. Different tools can be used to retrieve these data.

Honeypot is divided into two types: honeyd and Kfsensor. Honeyd is exclusively available for Linux users. Kfsensor is only compatible with Windows systems. Kfsensor communicates with many forms of malware and attackers via a variety of ports. Honeyd creates a virtual honeypot for attackers to interact with. After examining both honeypots, we discovered that Honeyd has detected a large number of suspicious IP addresses. However, Kfsensor displays information on the IP address that they have assigned [19]. In the field of network security, this procedure is implied.

A honeypot is difficult to set up in general since there is little knowledge on the subject and architectural ideas are kept secret to prevent attackers from finding and circumventing such systems. The honeypot should be difficult for automated tools to exploit, but it should also be easy to find, and we must ensure that there is a human attacker lurking in the background. In addition to intrusion detection systems and

firewalls, honeypot sites provide a complimentary answer to what is currently available in terms of security[20].

A. SETUP TOOLS

Without an operating system, a virtual machine running the Honeypot has direct access to the hardware it is running on and the virtual machines running on top of the hypervisor. We can use this to run some tools directly. To do so, there is a need to set up some tools in the honeypot. We can track down the attackers who are targeting our IP address using a variety of methods.

Most organizations and government agencies use honeypots on their networks because honeypots are a better solution for obtaining information about attackers. A honeypot looks like a real computer system. It is basically the same as a real type of server, but it's a trap server. The browser store is a place where one can hide the information.

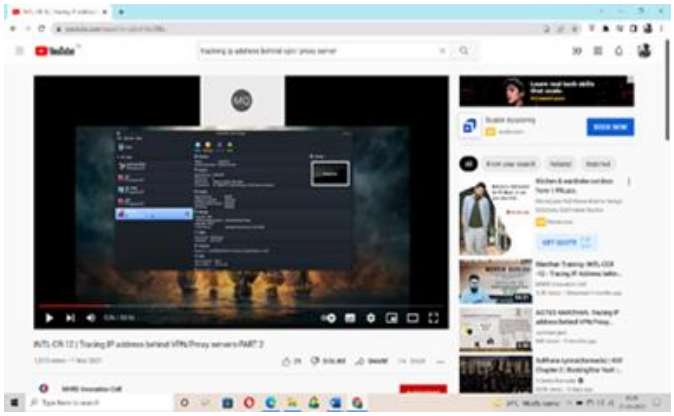


Figure. HoneyPot

Figure. Start Terminal

B. ngrock

ngrock is a port forwarding service. This tool will demonstrate what the ng rock is and how it functions on the ng raw, so go to the ng rock website and download the ng rock and ng rock all right. Simply get the ng rock from the official np rock website. The ng rock is about forwarding local systems to the you know online internet server, so this is our linux base.

On here, you can simply sign up with the Ngrok and then download and then create your authentication token and then set up your authentication and you know start the http server which is your http server put your local tool publicly right let me show you but let's just finish on this and the second thing which is the juicy information or the juicy file and then payload.

C. Canary token

The canary token is good for uh collecting information from the attacker right so it's basically the canary token. You know the behind the Ip and you know the actual location right so that's a very reliable the canary token the canary token is good for uh collecting information from the attacker right so it's basically the canary token.

Let me just go to the canary token right here it open make it open right so the canary token right so this is the official website and also there is a github you know page for the track activity action for the other network right so there are so many payloads related to the collecting information and there is it all right so some images if that collapse if that click on that image or automatically detect it all right so some images if that collapse if that click on that image or automatically detect it all right so some images.

If you collect the IP address on any file, you may simply generate it. There are all these tokens on here, but we're working on the xml file, so just try to put the xml file on my honeypot server and then you try to that, so this is a simple way all right you can just go

on there is just xml file so when the attacker opens the xml file in the system, it automatically detects the Ip address of the real-life address right when I just click on it and let's create a fake mail.

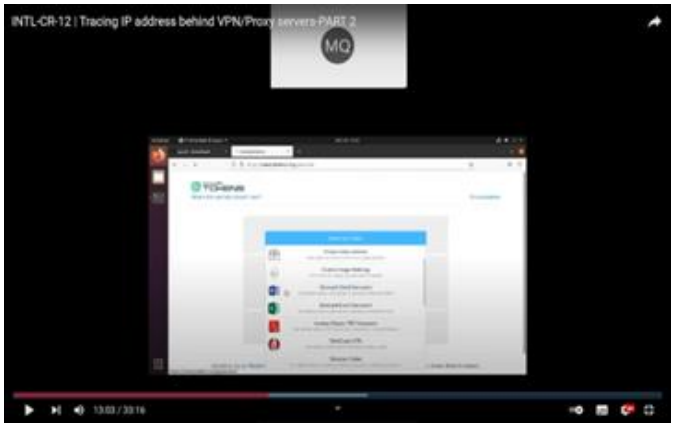


Figure. Canary Tokens

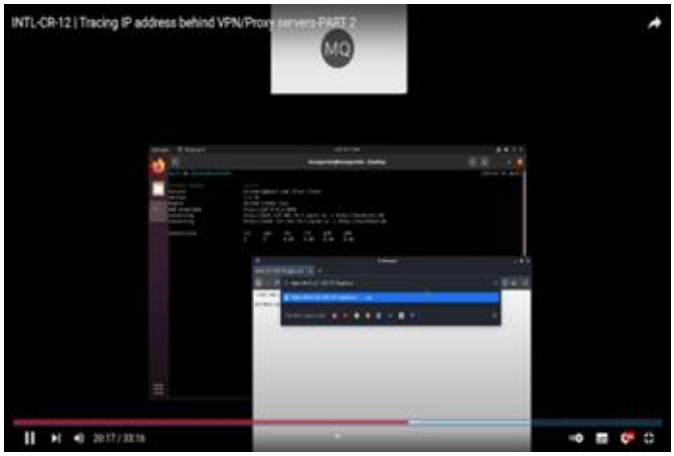


Figure. Ngrok http 80

This reveals the attacker's mac address, which, when entered into the Tor browser, yields one IP address, which is then used to create a leak DNS, revealing the attacker's address and the place where he is located.

Information which is the attacker is you know trigger attacker is a fish right so when you download any files from the server so when there's open this actually the real you know the real systems automatically detect the attacker actually the side you know what the ipad does is it's running on that right but the live right better and there are so many options on there at the barcode and SQL server and clone the website so that attack isn't it' so this is happened in this attacking.

IV. CONCLUSION

The need to develop and research solutions is being driven by the large number of attacks and exploitations of vulnerabilities. This survey presents a solution for tracing IP addresses behind VPN/proxy based on identifying attackers, redirecting them to a controlled environment, and analyzing their actions. Many prominent web applications still run with known and unsolved vulnerabilities, necessitating the development of solutions to ensure web application protection systems. consist of data and statistics gathered from captured attacks, revealing insights on an attacker's tools and activity.

We may evaluate and determine the competency level of the attackers that are attempting to break the honeypot by starting with an acceptable environment and a honeypot built on a Capture the Flag game with several levels of difficulty. They opened the excel page with their knowledge, and our code saved the information on them at that time. We can easily manage the attacker's information with this type.

V. REFERENCES

[1]. Miller, Shane & Curran, Kevin &Lunney, Tom. (2020). Detection of Virtual Private Network Traffic Using Machine Learning. International Journal of Wireless Networks and Broadband Technologies. 9. 60-80. 10.4018/IJWNBT.2020070104.

[2]. Pavlicek, A &Sudzina, F 2018, 'Internet Security and Privacy in VPN', Journal of Networking Technology, vol. 9, no. 4, pp. 133-139. https://doi.org/10.6025/jnt/2018/9/4/133-139.

[3]. X. Mi et al., "Resident Evil: Understanding Residential IP Proxy as a Dark Service," 2019 IEEE Symposium on Security and Privacy (SP), 2019, pp. 1185-1201, doi: 10.1109/SP.2019.00011.

[4]. S. -H. S. Huang and Z. Cao, "Detecting Malicious Users Behind Circuit-Based Anonymity Networks," in IEEE Access, vol. 8, pp. 208610-208622, 2020, doi: 10.1109/ACCESS.2020.3038141.

[5]. McKeague, Jonathan & Curran, Kevin. (2018). Detecting the Use of Anonymous Proxies. International Journal of Digital Crime and Forensics. 10. 74-94. 10.4018/IJDCF.2018040105.

[6]. M.Pannu, B. Gill, R. Bird, K. Yang and B. Farrel, "Exploring proxy detection methodology," 2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF), 2016, pp. 1-6, doi: 10.1109/ICCCF.2016.7740438.

[7]. Yogita Chhabra Assistant Professor, Department of Information Technology, Ajay Kumar Garg Engineering College, Ghaziabad -A STUDY OF RECENT RESEARCH TRENDS OF PROXY SERVER in May 2015.

[8]. Y. Chang, K. Yoon and D. Park, "A Study on the IP Spoofing Attack through Proxy Server and Defense Thereof," 2013 International Conference on Information Science and Applications (ICISA), 2013, pp. 1-3, doi: 10.1109/ICISA.2013.6579417.

[9]. Vikas Mishra, Pierre Laperdrix, Antoine Vastel, Walter Rudametkin, Romain Rouvoy, Martin Lopatka-Don't count me out: On the relevance of IP addresses in the tracking ecosystem in jan 2020 https://doi.org/10.1145/3366423.3380161.

[10]. James Williams IMPERIAL COLLEGE LONDON DEPARTMENT OF COMPUTING-Identification of IP addresses using fraudulent geolocation data in June 2020.

[11]. A.Tosun, M. De Donno, N. Dragoni and X. Fafoutis, "RESIP Host Detection: Identification of Malicious Residential IP Proxy Flows," 2021 IEEE International Conference on Consumer Electronics (ICCE), 2021, pp. 1-6, doi: 10.1109/ICCE50685.2021.9427688.

[12]. Vijayakumar, K. & Rai, Achyut & Kumar, G.Senthil& Angel T.S, Shiny &Snehalatha, N.. (2020). A two-way approach for detection and prevention of IP spoofing attacks. AIP Conference Proceedings. 2277. 020002. 10.1063/5.0025453.

[13]. Wang, Z., Li, H., Li, Q. et al. Towards IP geolocation with intermediate routers based on topology discovery. Cybersecur 2, 13 (2019). https://doi.org/10.1186/s42400-019-0030-2

[14]. L.Su, D. M. Divakaran and V. Thing, "Privacy preserving IP traceback," 2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA), 2018, pp. 1-8, doi: 10.1109/ISBA.2018.8311469.

[15]. Englehardt, Steven & Narayanan, Arvind. (2016). Online Tracking: A 1-million-site Measurement and Analysis. 1388-1401. 10.1145/2976749.2978313.

[16]. Wang, Yulin&Xiong, Dongliang. (2010). IP tracking for DoS/DDoS - State of the art. International Conference on Internet Technology and Applications, ITAP 2010 - Proceedings. 10.1109/ITAPP.2010.5566237.

[17]. Perino, Diego &Varvello, Matteo &Soriente, Claudio. (2018). ProxyTorrent: Untangling the Free HTTP(S) Proxy Ecosystem. 197-206. 10.1145/3178876.3186086.

[18]. Y. D. Mane and U. P. Khot, "A Systematic Way to Implement Private Tor Network with Trusted Middle Node," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-6, doi: 10.1109/INCET49848.2020.9154076.

[19]. N. Bhagat and B. Arora, "Intrusion Detection Using Honeypots," 2018 Fifth International Conference on Parallel, Distributed and Grid

Computing (PDGC), 2018, pp. 412-417, doi: 10.1109/PDGC.2018.8745761.

[20].O. Grigorescu, C. Săndescu and A. Caba, "Web Application Honeypot Published in the Wild," 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), 2020, pp. 1-6, doi: 10.1109/RoEduNet51892.2020.9324870.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Survey on Gray Hole Attacks Prevention Mechanisms

Subash Chaudhary, Manikandakumar M, Mahantesh K
Department of CSE, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

There are many types of Ad-hoc networks, so security is the most important concern in different networks like IoT, MANETs, VANETs, etc., These are all types of networks the attacks can happen maybe gray hole attacks or black hole attack. So, this paper will survey the gray hole and black hole attack. In today's scenario in Ad-hoc networks, it is used for many applications which are used in real-time like disaster management, military surveillance, etc. In These types of network, there will be the possibility of information leakage. Selective drop-in packets will occur in a gray hole attack. Due to packet loss, the information can't be transmitted. The malicious node will act as the destination node so that the packet loss will take place and it is known as the attack over the network. This study paper presents some findings on the issues of gray hole attacks and also suggests some prevention methods to reduce or avoid the attack.

Keywords: Ad-hoc Network, Gray hole, Blackhole, MANET, VANET

I. INTRODUCTION

Ad-hoc networks assign groups whose specialty scatters and committed nodes for watching and recording the condition of the physical environment which can be temperature, humidity, sound, etc. Manets can also be used for organizing data that are collected at a central location. Some dependent infrastructure which is wired communication is linked, which are used for creating the connections to nodes.

VANETs is a self-organized network that can be created and linked by traveling through the road to Infrastructure like lots of parking. This network is being used in order to improve driving safety. Whichthe driver needs and it also supports traffic management. It enables the connection between units that are of two types. One is a pure wireless network and the other is that with fixed infrastructure.

There are five sections to this study. The different types of networks are discussed in the first section the Gray-hole attack can occur. The second section II will summarise existing research on Gray-hole attacks and other attacks like black hole attacks, and worm whole attacks in the networks. The III section is divided into the research technique employed in this study, as well as the instruments employed and the recommended layout. The research results are explained in Section IV, and in section V, there is a conclusion of all the paper's future work.

II. EXISTING WORK

In Paper [1] in this paper, their research work is for IoMT devices based on the 5G network. due to IoT malware developed which becomes an issue, this malware has to potential to disrupt IoMT connections as well as it can also harm the medical smart devices. While initialization of node there will be a unique key shared between nodes. That key will validate the transmission of the packet by improving network security, the packet to the next-hop node so the purposed system will exhibit performance analyst by 98.23% packet delivery ratio as well as end to end 10% of transaction delay which provides the outperforming performance.

In Paper [2] The network layer attack will dominate the routing capacity of AODV protocols. Sometimes The grey hole attack causes some or all data packets from that node to be dropped. which will behave like a single Gray hole attack, some of the nodes generate fake RREP to convince the destination node. So, a security mechanism is required for important or sensitive and confidential data. In Paper [3] the comparison of the network performance of MANET-IoT under various attacks, this paper has implemented attacks by modifying NS-3 to AODV routing protocols.

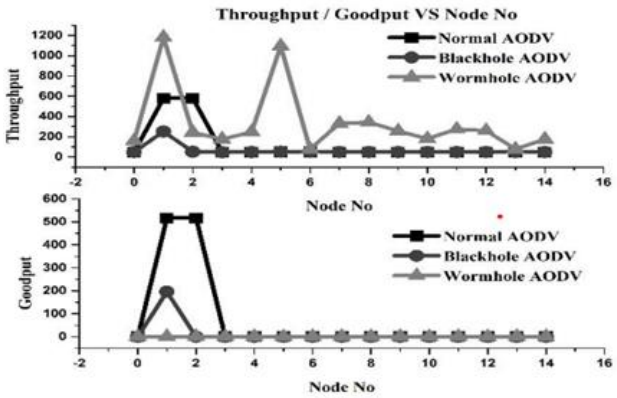


Fig 1: comparison between goodput/throughput Vs Node

In Paper [4] Under the real-world experiment, conditions We need to assess the impact of GH attacks using PMU data. The electrochemical modes are between 0.3 and 0.45 Hz, and for various Gray magnitudes, a simulated attack is carried out. The packet drops rates up to 20%, this model obtains the LS algorithm if there are Gray hole attacks 40 % of magnitude has negligible impact. According to the model estimate algorithm, an investigation was conducted to assess the effects of various packet drop models for the Gray-hole attack.

In Paper [5] the algorithm has been implemented based on their security methodsthe statical-based Gray-hole attack algorithm is used for the ratio of the transmission efficiency for finding the malicious node the exception finding algo is used. the algorithm is used to find the percentage of the packets that were discarded as a result of locating the rogue node

In Paper [6] the comparison between exiting work with new has been done by using the AI concept the model is developed to detect the intrusion automatically in the network which will achieve PDR's average score of 97.55 percent. Using this technology, intermediate nodes deliver data packets without delay to another node.

In Paper [7] in sensor nodes, a cryptographic solution can't be implemented. According to this limitation for many security attacks, WSN is opened. the black ole the performance of the network is harmed as a result of the attack.

In Paper [8] to combat the various sorts of MANET attacks, The approach of malicious node detection was applied, as well as encryption. The attacker node is detected via the reverse tracking method.

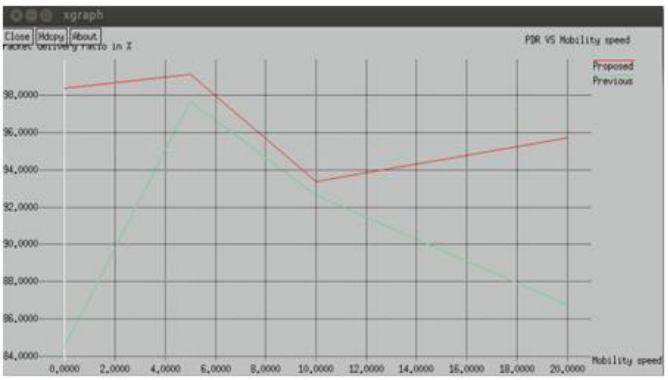


Fig 2: PDR vs mobility speed

In Paper [9] Multiple connections were successfully mitigated against black holes and grey holes. 70% to 80% of packet delivery ratio has been achieved and the dropping ratio of packets is also decreased to 15%. But delays are increased from 0.2 to 0.5 sec.

In Paper [10] detection mechanism has purposed for a Gray hole attack in DSR protocol in that the aggregated signature algorithm is used also the check-up and diagnostic is used for each node so that they can produce forwarding packets evidence, to trace the malicious nodes. The local co-operative detection technique to identify malicious nodes in a Gray-hole attack is suggested.

In Paper [11] the proposed can minimize the Gray-hole attack by using BFST and MST which are used to identify the attacker which is the basic idea of contradiction in between the nodes then the message can be sent from source to destination. In this paper, they are finding the presence of the attacker if the attacker is present immediately the dropping of paths happens.

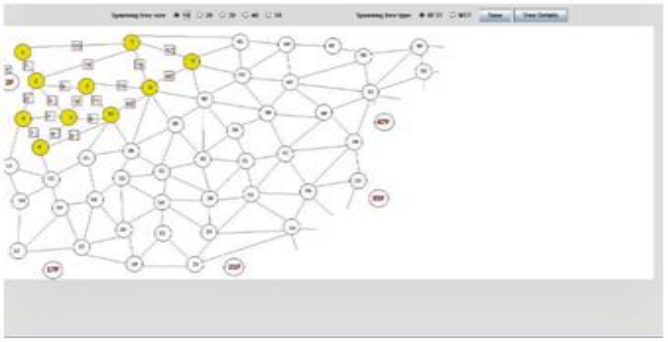


Fig 3. Topology setting

In Paper [12] the purposed model is developed for resisting a smart black hole attack which is followed by a timer as well as by a baiting message the TTL is set to one to avoid the fake request if any node wants to communicate it will generate RREQ to the goal node. The path selection mechanism is developed so that the fake request for the shortest path can be avoided. From this model, the throughput and delivery ratio of the packet is also to decrease End-to-End delay.

In Paper [13] Selected drop attacks are effectively protected by RSDA. In a selective attack, adjacent nodes will be impersonated in order to transmit their messages to the next node, and the malicious node can also refuse to forward the message to another node.

In Paper [14] DCPM is used for the Gray-hole attack prevention is a function that contradicts itself which is developed for identification and also for the prevention of isolation attack of the node.

By dropping data, the attacker plans the attack which should be routed if they act or not as sole MPRs. The main end of DCPM is to avoid the selection of suspected bumps in the form of sole MPR which will help Gray-hole attacks. In Paper [15] the presented modules are also used to judge the performance network. The proposed model is also used for the isolation of the malicious nodes from the MANET

In Paper [16] This is used to analyze the grey hole attack, which is carried out using the ns-2.35 simulation. The simulation file is used to calculate the number of packets delivered, dropped, and received at varying node speeds.

III. RESEARCH METHODOLOGY

In Paper [17] for the detection of malicious plots, this paper is proposed. When the RRQ is received from other nodes, the devices will validate the presence of a malicious node and begin reversible tracing to locate

the offending node and prohibit further routing. It will move with the help of two one-hope neighbors.

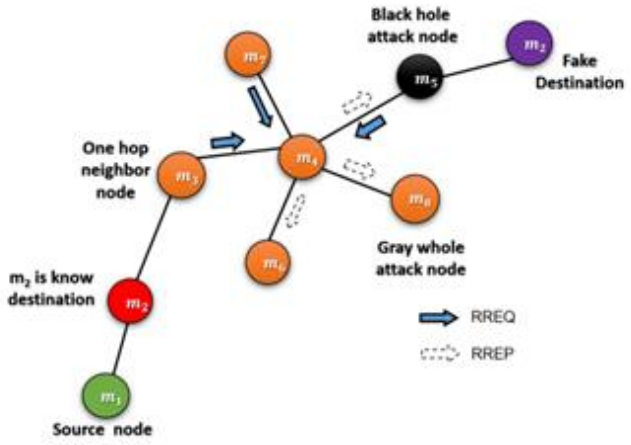


Fig 4: Detection of dual-cooperative bait model

In Paper [18] comparative bait system has been proposed to detect the collaborative gray-hole or black hole attacks. Using AODV this approach has been implemented to reduce the routing overhead. In the simulation, it performs better than using CBDS in DSR. This proposed system is to detect attacks. The result shows the increased ratio of packet delivery. It also proposed an enhanced model for detecting and preventing the malicious nodes over the MANETs, in the single or collaborative attacks from grey and black holes The approach was also proposed utilizing data routing information protocols. Each node will be cross-checked by DIR.

In Paper [19] they have shown how multi-hop networks are adoptable in an IoT environment. Two nodes were employed, one for downstream traffic and the other for packet counting. The tracking that corresponds to coherent time is carried out during the interval. In a multi-hop wireless network, an algorithm has been suggested to detect a gray-hole attack. It is carried out Fig 5: Detection of dual-cooperative bait model of different algorithms during Gray hole attack and SGA. by a rational adversary as a Byzantine attack.

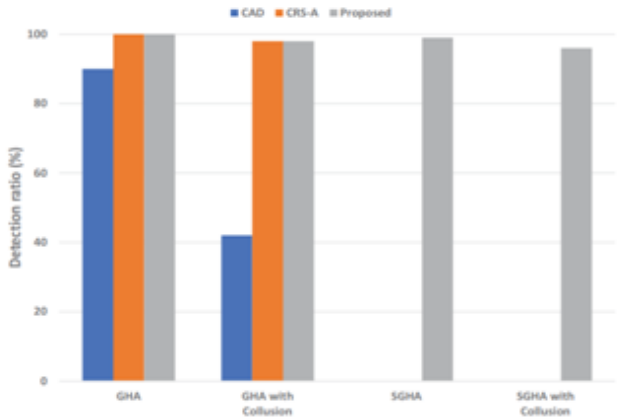


Fig 5: Comparison of Detection ratio

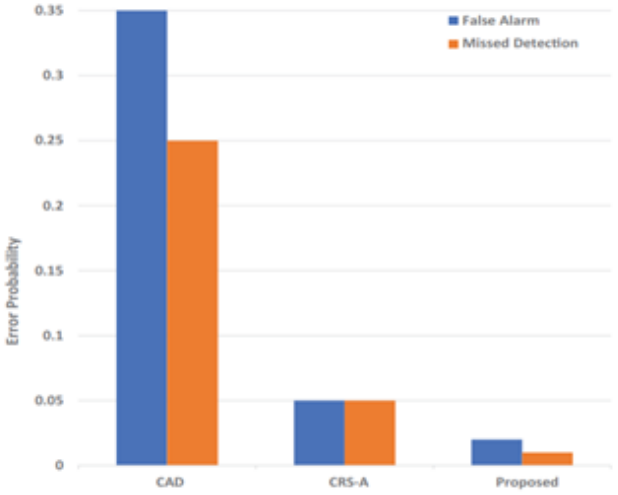


Fig 6: Comparison of error probability

Fig 6 is the comparison between the probability of error and how much It has detection rate and a lower mistake rate. While detection accuracy is a desirable quality in most security applications.

In Paper [20] They've devised a way to counteract grey hole attacks. Its goal is to monitor traffic and node behavior in order to identify malicious nodes. The IDS nodes are used in this strategy to preserve network efficiency. Gray hole assaults are prevented and detected in two phases by AGHA. 1. Detection during the Routing Discovery Phase 2. Detection during the Data Transmission Phase.

IV. SIMULATION SETUP

For simulation and implementation of the N-layer attack will be done by using in NetSimattacker.c file. To the workspace of the AODV project, the file is added which will generate a new AODV.dll which will be in the bin folder. This is shown below in figure 7.

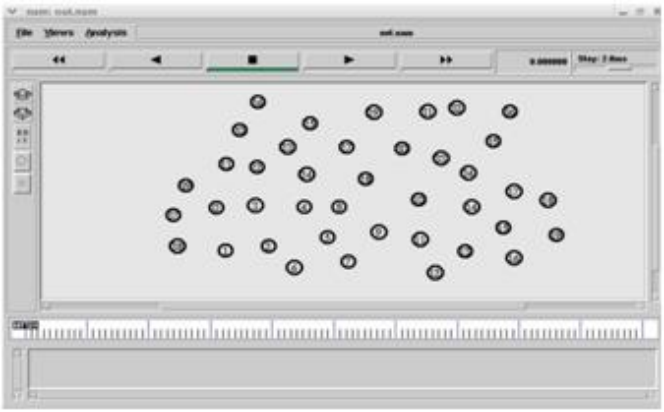


Figure 7: node initialization

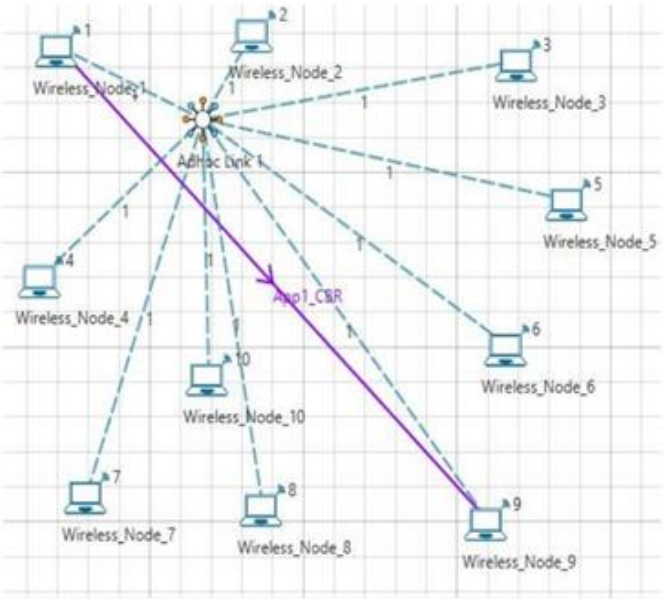


Fig 8: NetSim simulation set-up

From the fig 8, The node has been joined the wireless network, and the routing protocol with CBR traffic has been configured, as well as the Nam animation

V. RESULTS

Each of the forty nodes in the experiment was set to run for 10 seconds. It has a good network connection channel capacity, which is measured in bits per second (bit/s). When compared to existing systems, the proposed framework delivers more packets while avoiding attackers. Figure 9, shows the generation of nodes and detection of the malicious nodes by using the NS simulator so that the malicious nodes can be detected and it also shows the prevention of the malicious nodes to the networks from the attack.

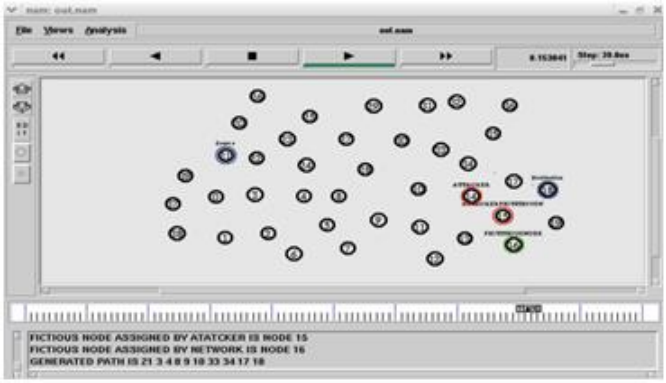


Fig 9: attack detection

VI. CONCLUSION

The demand for developing and studying remedies is being driven by a large number of attacks and exploits of vulnerabilities. By detecting rogue nodes utilizing a number of approaches and processes, this paper presents a strategy for reducing grey hole attacks on mobile ad-hoc networks and other networks. We looked at network attack detection and prevention in order to assure the security and safety of message delivery in this study.

We can analyze and detect the level of competence of attackers attempting to enter the network layer by creating a false request that pretends the node is the destination. It is increasingly probable in the future that detecting an attack over the network would

prompt the device to find the attacker, and if the attacker can be discovered, the message will travel faster from the source to the target node. It is also possible to turn the network into an anti-attack network. As the message moves from one node to the next, it will become more secure.

VII. REFERENCES

[1]. V. Balaji and P. Selvaraj (2022) "Gray-Hole Attack Minimization in IoMT with 5G Based D2D Networks" article in Tech Science Press DOI:10.32604

[2]. Ausaf Umar Khan, Milind Madhukar Mushrif, Manish Devendra Chawhan and Bhumika Neole "Performance Analysis of Adhoc On-demand Distance Vector Protocol under the influence of Black-Hole, Gray-Hole and Worm-Hole Attacks in Mobile Adhoc Network" (2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

[3]. Muhammad Nasir Siddiqui, Kaleem Razzaq Malik and Tauqeer Safdar Malik "Performance Analysis of Blackhole and Wormhole Attack in MANET Based IoT" (2021) International Conference on Digital Futures and Transformative Technologies (ICoDT2) IEEE | DOI: 10.1109/ICoDT252288.2021.9441515

[4]. Pal, Aditya Ashok, Siddharth Sridhar, James D. Follum and A. David McKinnonha"Impact of Gray-Hole Attacks on PMU Data-Based Mode Estimation" IEEE (2018) DOI: 10.1109/PESGM.2018.8585920

[5]. Krzysztof Stepień; AnetaPoniszewska-Marańda "Analysis of security methods in Vehicular Ad-Hoc Network against Worm Hole and Gray Hole attacks" IEEE (2020) DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00072

[6]. POOJA RANI, KAVITA, SAHIL VERMA and GIA NHU NGUYEN "Mitigation of Black Hole and Gray Hole Attack Using Swarm Inspired Algorithm with Artificial Neural Network" IEEE (2020) DOI: 10.1109/ACCESS.2020.3004692

[7]. Shoukat Ali, Dr.Muazzam A Khan, Jawad Ahmad, Asad W. Malik, Anis ur Rehman "Detection and Prevention of Black Hole Attacks in IOT & WSN" Third International Conference on Fog and Mobile Edge Computing (FMEC) IEEE (2018) DOI: 10.1109/FMEC.2018.8364068

[8]. Neha, Gurjot Kaur Walia, and Harminder Kaur "A Cipher Based Acknowledgement Approach for Securing MANETs by Mitigating Black and Gray Hole Attacks" ResearchGate (2018) July DOI:10.1007/978-981-13-0755-3_16

[9]. Sharma Hitesh Omprakash and Margam K. Suthar "Mitigation Technique for Blackhole Attack in Mobile Ad hoc Network" 2020 DOI: 10.1109/ICCCNT49239.2020.9225381

[10]. AthiraHarikrishnan, Jasmine Joseph, Fathima Manzoor and Tinu Thomas "Gray-Hole Attack Minimization for Ad-Hoc Networks Using Contradiction" (2019) DOI: https://doi.org/10.30534/ijiscs/2019/19822019

[11]. Niranjan Panda & Binod Kumar Pattanayak "Defense Against Co-Operative Black-hole Attack and Gray-hole Attack in MANET" Volume 2018. https://doi.org/10.1155/2018/9812135

[12]. Adwan Yasin and Mahmoud Abu Zant "Detecting and Isolating Drop Attack in Wireless Ad Hoc Networks" (2019) DOI: 10. Black-Hole Attacks in MANET Using Timer Based Baited Technique".Volume 2018.

[13]. T. Poongodi, Mohammed S. Khan, Rizwan Patan, Amir H. Gandomi, and Balamurugan

Balusamy “Robust Defense Scheme Against Selective 1109/ACCESS.2019.2896001

[14]. Ms. M. Krishna Prasanna and Dr. M. Neelakantappa “Minimization of Gray-Hole Attack Effect for Secured Mobile Ad hoc Networks” (2018) Volume 118 No. 24 2018 ISSN: 1314-3395(on-line version) url: http://www.acadpubl.eu/hub/

[15]. Zulfiqar Ali Zardari, Jingsha He, Nafei Zhu, Khalid Hussain Mohammadani, Muhammad Salman Pathan, Muhammad Iftikhar and Muhammad Qasim Memon “A Dual Attack Detection Technique to Identify Black and Gray Hole Attacks Using an Intrusion Detection System and a Connected Dominating Set in MANETs” (2019) DOI:10.3390/fi11030061

[16]. MoirangthemMarjit Singh and Jyotsna Kumar Mandal “Gray Hole Attack Analysis in AODV Based Mobile Adhoc Network with Reliability Metric” DOI: 10.1109/CCOMS.2019.8821671

[17]. Osamah Ibrahim Khalaf, F. Ajesh, Abdulsattar Abdullah Hamad, Gia Nhu Nguyen, and Dac-Nhuong le “Efficient Dual-Cooperative Bait Detection Scheme for Collaborative Attackers on Mobile Ad-Hoc Networks” (2020) DOI: 10.1109/ACCESS.2020.3045004

[18]. Oluwafemi Fasunlade, Shikun Zhou, David Adrian Sanders " Comprehensive Review of Collaborative Network Attacks in MANET” Conference: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) DOI:10.1109/COMPSAC48688.2020.00-36

[19]. Anjana Prabhakar, Tricha Anjali “Gray Hole Attack as a Byzantine Attack in a Wireless Multi-Hop Network” (2019) DOI: 10.1080/19361610.2019.1669988

[20]. Omid Moradipour, Mohammad Fathi “An AntiGray Hole Attack Scheme in Mobile Ad Hoc Network” (2020)DOI: https://doi.org/10.1007/s10776-020-00492-6

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Recommendation and Simulation of Medical Diagnostics Using Artificial Intelligence

Dhanush Y P, Muskan Rizwan Shaikh, Yajat Vinod Vishwakarma, Surya Pandey
Department of Computer Science and Engineering, New Horizon College of Engineering, Marathalli, Bangalore, Karnataka, India

ABSTRACT

The healthcare industry is unique in comparison to other industries. It is a high-priority industry, and consumers want the highest level of care and services possible, regardless of the costs involved. Along with its subjectivity, the complexity of the condition, and the wide variations that exist across diverse interpretations, a medical diagnosis by human professionals is severely limited. This paper proposes an architecture for recommendation and stimulation of medical diagnostics in order to assist doctors and other medical professionals in the correct interpretation of a disease. The methodology followed uses the latest artificial intelligence and machine learning models to get precise results. To strengthen the architecture, we have made use of unsupervised learning, clustering and augmentation approaches. The datasets that will be used are imagery data, historic medical records, numeric data charts that describe various data points.

Keywords—Machine Learning, Unsupervised Learning, Clustering, Data Augmentation, Medical Diagnostics

I. INTRODUCTION

The science of diagnosis is a difficult task, as it requires the human mind to understand and connect all the information required to solve a problem. In medical diagnosis, the margin for error is zero. Throughout all of human history, we have questioned the very fabric of nature and reality, yet our accuracy in understanding how things work is still not efficient.

Technological advancements over the decades have come with a great promise, however their slow adoption in the medical field is quite surprising. A modern system which has the processing power, and the abundance of data should have paved the way for artificial intelligence to be used in almost every aspect in the field of medicine. But artificial intelligence applications are faced with some challenges, including the black-box nature of some AI models. The poor explainability of these black-box models leads to distrust from medical experts to make explainable clinical inferences [1]. There are often millions of parameters in Deep Learning models, and they only return a final decision result without any explanation. Due to the lack of transparency of deep neural networks, it is hard for the user to judge whether the decision is reliable, compromising trust with doctors.

Algorithms used to identify diseases rely on associative inference. This means that they identify diseases based on how strongly they correlate with a

patient's symptoms and medical history. In contrast, doctors perform diagnosis by selecting diseases that offer the best causal explanations for their patients' symptoms.

The application of artificial intelligence within the diagnostic process supporting medical specialists could be of great value for the healthcare sector and the overall patients' well-being. The integration of artificial intelligence into existing technical infrastructure accelerates the identification of relevant medical data from multiple sources which are tailored to the needs of the patient and the treatment process. Furthermore, artificial intelligence generates results based on a larger population rather than on subjective, personal experiences and achieves equal results when using identical medical data and does not rely upon situations, emotions, or time of day.

In Section 2, we give a literature review which enabled us to write this paper. Section 3 will give a detailed explanation of the proposed system and its models for the analysis of a case, which will make the process reliable and precise. Section 4 includes the results and discussions and finally Section 5 concludes the findings of the research paper.

II. LITERATURE REVIEW

The technological landscape has drastically evolved in the last few decades. This has propelled the development of new tools and improvement in our approach to solve problems using computation.

Puneet Mishra et al. [2] states that the goal of data preprocessing is to remove unwanted variability so that the useful information related to the property of interest can be used for efficient modeling.

A.D.Dongare et al. [3] interpret ANNs as massive parallel computational models that imitate the function of the human brain. A single computing node alone is not a very powerful computational engine. A single computing node generates a scalar output, which is a simple non-linear function of its inputs.

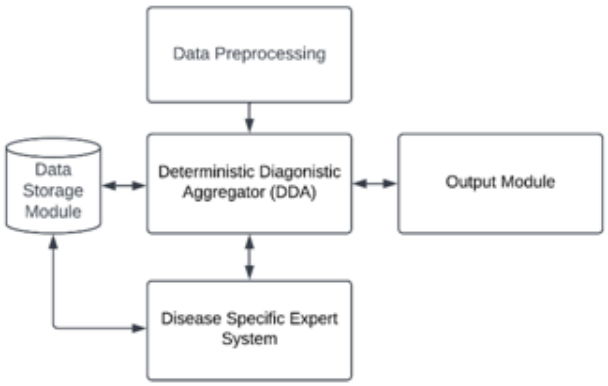
The use of both smart data augmentation and ANN make it possible to tackle the challenging nature of performing diagnostics in the medical field.

III. METHODOLOGY

A. Architecture Description

The fundamental approach used to solve a problem of such magnitude often requires the usage of various components working hand in hand. Every intelligent decision- making system uses a systematic workflow of how data moves around to provide accurate results. In this architecture which is shown in fig. 1, we divide the entire decision-making procedure into modular blocks, each of which provides a functional value.

Data is first collected and filtered which is then passed to a decision-making block that constantly communicates with all the diagnostics machine learning models. The decision- making block is the central nervous system of the entire architecture. It constantly keeps updating its data repositories and maintaining a dialogue between all the individual diagnostic modules. The architecture is backed by a persistent storage module which maintains all the processed data. Finally, the end result is validated through a feedback loop which initiates all the modules to calibrate.

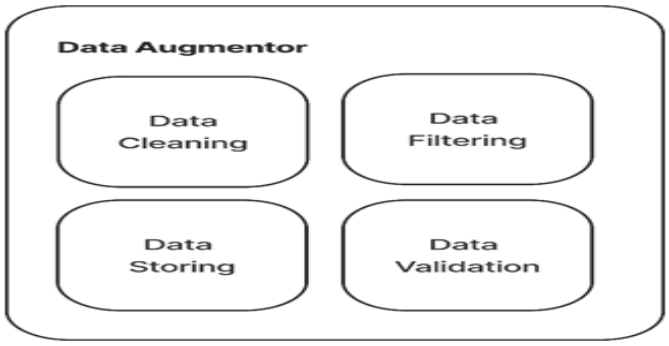


Proposed Architecture

- Data Preprocessing

Data is everything when it comes to the process of making decisions. There is an abundance of data and with this privilege comes the necessity to classify them into meaningful categories. Data can be of different types like, images from X-Ray, MRI etc., historical medical records, numeric data charts describing various data points. In contemporary systems, data from patients is more complex than a simple table of values. Complex data types, such as time series data, are used to describe attributes within the patient information. For example, it is useful in medicine to compare patient examinations at two different time points in order to discover the progress of the process of recovery from a disease [4].

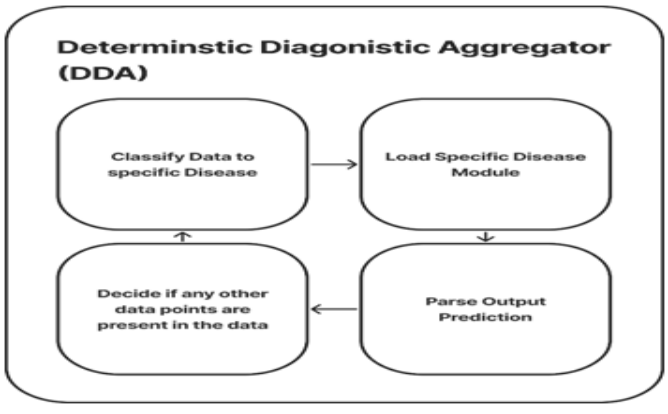
Data noise, which can negatively affect the quality of imaging outputs, is one of the major factors affecting biomedical image processing [5]. Hence all the unwanted properties and outliers need to be filtered. With the domain of the data being set, we now validate features that can be used in the workflow later.



Components of Data Preprocessing

B. Deterministic Diagnostic Aggregator

Even though the entire architecture is modular, there is a need for a centralized decision-making body which governs the working of all the individual modules. The Deterministic Diagnostic Aggregator (DDA) plays this role by firstly receiving data from the preprocessor and performing pre diagnostic tests to understand the nature of the data with which it then maps to the particular Machine Learning (ML) model which is responsible for analyzing the data points and providing the results. These results are then sent back to the DDA which then verifies it for accuracy and unexplored patterns. DDA itself could be a ML model to classify which disease can be detected in the input data. The DDA reiterates its results and is in a feedback loop that constantly tries to find correlation and patterns in the data [6]. The working of DDA is depicted in Figure 3.

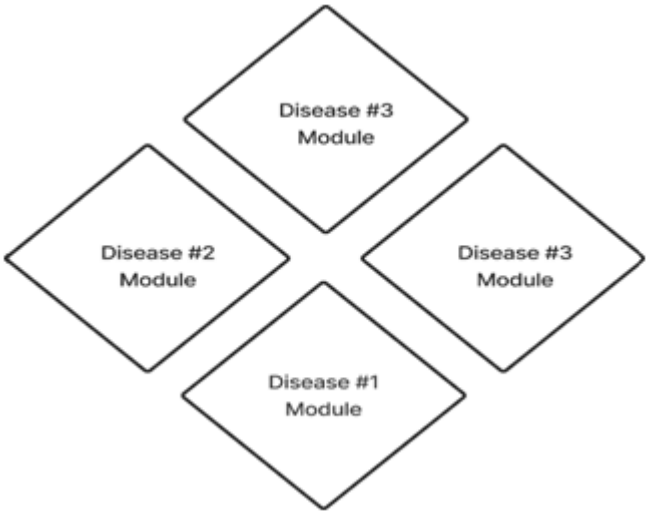


Internal functioning of Deterministic Diagnostic Aggregator

For instance, if an X-Ray for a test of bronchitis is input to the system, DDA will be able to not only detect if the input data suffers from bronchitis but also detect if any other anomaly exists in the X-Ray. The DDA is tightly coupled with the Disease Specific Expert System (DSES). A more detailed explanation of the DSES can be found in part D of this section.

C. Disease Specific Expert System

Every disease has its own unique identification factors that cannot be generalized and applied to reach a good diagnosis. The use of the DDA which acts as an interface to enable the offloading of decision making to the machine learning model which is best suited allows the existence of models that are specific to a particular disease.



Modular DSES

Each ML model has the ability to stay isolated from all other decision-making entities which allows them to be unbiased. Having such an abstraction creates opportunities to enhance only parts of the decision-making system without affecting the working of the entire system. During the training procedure the models, the data that is only relevant for that specific disease diagnosis is loaded by the DDA. The expert ML models upon training provides a summary of all the parameters that were considered and the result back to the DDA.

D. Decision Output

The thinking process behind the way we interpret a decision is only unique to our ability as humans because of our advanced linguistic capabilities. Hence if we are to create a recommendation system for diagnostic purposes based on algorithms then we are bound to enable better interpretations of the results without any loss of data.

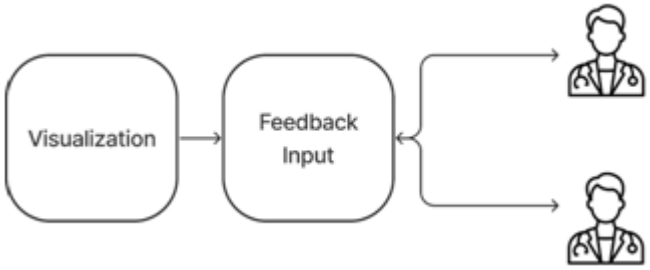


Fig:Output Module, Processes input from DDA for visualization.

This module enables us to convert algorithmically generated data into human understandable explanations [7]. With the ability to form conclusions which we cannot fathom as to how they were made; the output must include a traceback that includes reasoning for each step in the decision-making process. The output module is the interface between the architecture and the doctors. It also functions as the input to the feedback loop which actively fine tunes the specific parts of the architecture. By doing this the model is in a state of continuous innovation and is actively learning new relations in the data. As depicted in Fig. 5 the feedback loop can help optimize separate expert systems and the DDA to improve accuracy over all.

E. Data Store Module

This module is responsible for systematically storing all the inputs in an organized manner for quick access

by the DSES and other components of the system. The output data of the DDA is also stored for future reference and to implement transfer learning when enough data points exist [8].

The Data Store Module can hold image data using a file system, relational data using RDBMS and unstructured data using NoSQL database as per requirements. Fig. 6 gives an overview of the Data Store Module module.

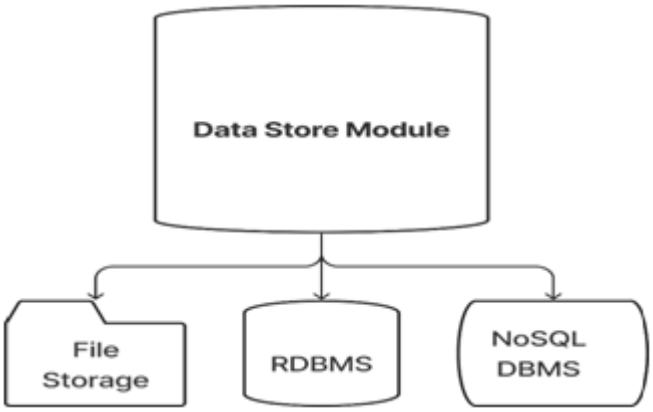


Fig:Data Store Module

IV. RESULTS AND DISCUSSION

Through our findings and research, we have built a precise and relevant architecture that can be implemented in any environment easily. Our architecture includes the human AI interaction that plays an essential part in the analysis and is the key point in achieving a model that is not only dependent on the dataset but also accepts real time inputs [9].

In addition to the points mentioned above our architecture is modular in nature and will work well on various categories of data, which makes it versatile.

V. CONCLUSION

AI technology has proven to be the most advanced innovation of humankind. It has taken over various sectors, it is about time that we leverage its functionalities to a field– medical, that is finding it hard to adapt to the latest technologies in the diagnosis sector. Diagnosis of a disease using AI involves various inputs and analysis which might overwhelm an individual but with the approach provided in this architecture, we can encourage more reliability in the system [10].

VI.REFERENCES

[1]. Amato, Filippo, et al. "Artificial neural networks in medical diagnosis." Journal of applied biomedicine 11.2 (2013): 47-58.

[2]. Mishra, Puneet, et al. "New data preprocessing trends based on ensemble of multiple preprocessing techniques." TrAC Trends in Analytical Chemistry 132 (2020): 116045.

[3]. Dongare, A. D., R. R. Kharde, and Amit D. Kachare. "Introduction to artificial neural network." International Journal of Engineering and Innovative Technology (IJEIT) 2.1 (2012): 189-194.

[4]. Aljawarneh, Shadi, et al. "Particularities of data mining in medicine: lessons learned from patient medical time series data analysis." EURASIP Journal on Wireless Communications and Networking 2019.1 (2019): 1-29.

[5]. Wang, Shuihua, et al. "Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects." Information Fusion 76 (2021): 376-421.

[6]. Blasch, Erik, et al. "Methods of AI for multimodal sensing and action for complex situations." AI Magazine 40.4 (2019): 50-65.

[7]. Adadi, Amina, and Mohammed Berrada. "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)." IEEE access 6 (2018): 52138-52160.

[8]. Yang, Yahan, et al. "Standardization of collection, storage, annotation, and management of data related to medical artificial intelligence." Intelligent Medicine (2021).

[9]. Jarrahi, Mohammad Hossein. "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making." Business horizons 61.4 (2018): 577-586.

[10]. Johnson, Kevin B., et al. "Precision medicine, AI, and the future of personalized health care." Clinical and translational science 14.1 (2021): 86-93.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Performance Evaluation of GPSR for Wireless Network Systems

Thanu V, Manikandakumar M, Asha Rani Borah, Jeevitha R
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Greedy Perimeter Stateless Routing (GPSR) is a position based wireless network routing protocol. The data forwarding in GPSR takes place in the form of Greedy forwarding and Perimeter forwarding. Greedy forwarding by using information about a router's immediate neighbors in the network topology. When greedy forwarding fails, the algorithm performs routing around the perimeter of the region. As locations frequently changes, GPSR can route data efficiently by finding correct new routes quickly. This paper gives the Performance evaluation of GPSR protocol based on following metrics: Packet Delivery Success Rate (DSR), Routing Protocol Overload (RPO) and Path length. We have shown the performance of GPSR on wireless networks to that of Dynamic Source Routing.

Keywords: Wireless Networks, geographic routing, GPSR, DSR

I. INTRODUCTION

In Wired networks, there is a standard use of Distance Vector (DV), Link State (LS) and Path Vector routing algorithms to provide scalability. Whereas in Wireless networks which has an increased mobility, the problem of finding a correct route from source to packet destination is challenging. When nodes are changing from one place to another, the topology of the network can change rapidly. Such networks require a routing protocol that finds a correct path even as nodes changes.

GPSR is a position-based routing protocol based on geographic position information while other routing algorithms are graph-based notions of finding the shortest paths and to find routes. We propose the use of GPSR to achieve scalability under increasing number of nodes and increased mobility rate.

GPSR allows networks which uses a large number of nodes or mobility to achieve scalability include:

- As-hoc networks: these networks have no fixed infrastructure, used for business conference or for military uses.
- Sensor networks: are mobile networks with large number of nodes and have impoverished per-node resources.
- Rooftop networks: it is fixed network and deployed densely in cities with vast number of nodes.

We will show that GPSR allows routers to be stateless and that it can find correct path by making Greedy forwarding decisions, when it fails then the algorithm recovers by doing Perimeter forwarding.

This paper is going to be devised according to the following: Section 2 describes algorithms that comprise GPSR. Section 3 presents the performance of GPSR in simulated mobile networks and conclude by summarizing our findings.

II. GPSR ALGORITHMS

The Greedy Perimeter Stateless Routing (GPSR) protocol is a novel routing protocol for wireless networks. GPSR algorithm consists of two methods for forwarding packets: greedy forwarding, which is used wherever possible, and perimeter forwarding, which is used in the regions where greedy forwarding fails.

A. Greedy forwarding

In GPSR, sources send packets marked with their destination locations. As a result, if a forwarding node knows its immediate neighbour's position, then it can make a locally optimal choice of next hop to the packets destination. A forwarding node between source and destination receives the route request packet then it checks its table to send the packet to the node nearest to the destination node and this takes place until the destination is reached.

Fig 1: shows an example of Greedy Forwarding. Here x receives a packet from source destined to D. x forwards the packet to y as y is the closest neighbour to D and this continues until the packet reaches destination.

When the packet arrives on a node where the neighbours position can't be found and greedy forwarding won't take place then a second algorithm called Perimeter forwarding will be used. In GPSR protocol, all nodes of the network have a local table which consist of neighbour nodes position. The table is refreshed on a regular time interval. The packet sent by the source with its destination address would not have its address changed by node which forwards the packet to destination. The following table shows the header fields:

TABLE 1. THE GPSR HEADER FIELDS

Field	*Function*
D	Destination Location
Lp	Location Packet Entered Perimeter Mode
Lf	First node encountered in the face of the planar graph
e0	First Edge Traversed on Current Face
M	Packet Mode: Greedy or Perimeter

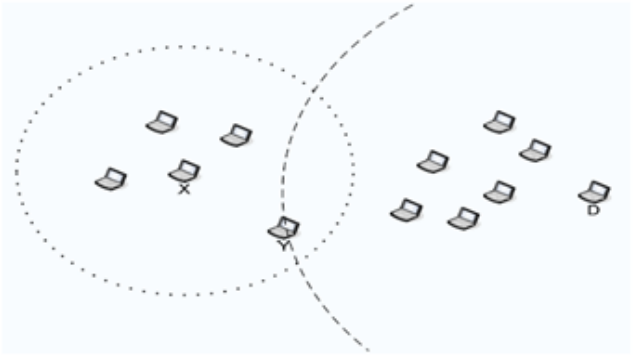


Figure 1. Greedy forwarding example. Y is X's closest neighbor to D.

B. Perimeter forwarding

When a node can't find its immediate neighbour node nearest to the destination with the Greedy forwarding algorithm then the packet mode will be changed into Perimeter mode. In fig 2: although there are 2 paths, (x-y-z-D) and (x-w-v-D), exist to D, x will not send the packet to w or y using greedy forwarding. At this point the algorithm recovers by using perimeter forwarding.

The Perimeter algorithm tries to send the packet along the empty region known as void to a node nearer to the destination. The packet arrives on the node nearer to destination after being forwarded on the edge of the empty plane. First the node will check if any one of its edges lies on the vector XD(fig.2) the header will store the position of this intersection on position Lf. With this information, the algorithm tries

to find the next edge and plane to send the packet, In order to prevent a packet to pass endlessly on the edge of a plane, the header flag will be set by reaching the plane. The packet will change back in Greedy mode when the packet reaches a node closer to the destination in perimeter mode to position. Now the Greedy forwarding algorithm can be used again to forward the packet to the destination.

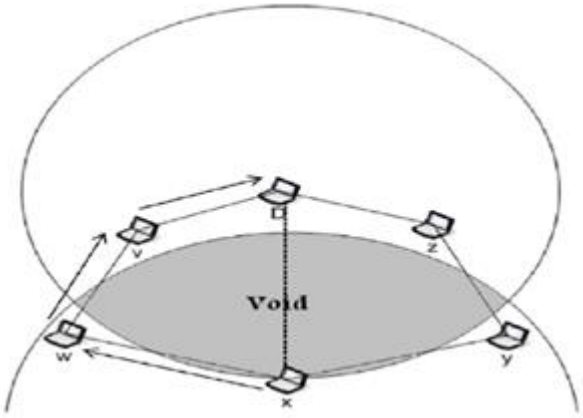


Figure 2. Perimeter forwarding

C. Planerized Graph

It is unacceptable for a routing algorithm to fail to find a route to a reachable node in a static network topology while the no-crossing heuristic empirically finds the vast majority of routes in randomly generated networks. So we present alternative methods for eliminating crossing links from the network.

A planar graph is one in which no two edges cross. A graph can be seen as a set of nodes with radios which are identical with circular range r. unit graphs are there whose edges are dictated by a threshold distance between vertices.

The two planar graphs commonly known are the Relative Neighborhood Graph (RNG) and Gabriel Graph(GG). A network with no crossing link can be formed from the RNG or GG when an algorithm removes the edges from the graph that are not part of the RNG or GG. In our application only information about the local topology is needed as the algorithms input.

The RNG is defined as An edge(u,v) exists between vertices u and v if the distance between them, d(u,v) is less than or equal to the distance between every other vertex w and whichever of u and v is farther from w.

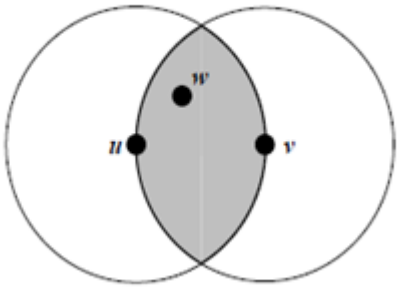


Figure 3: The RNG graph. For edge (u,v)) to be included, the shaded lane must contain no witness w.

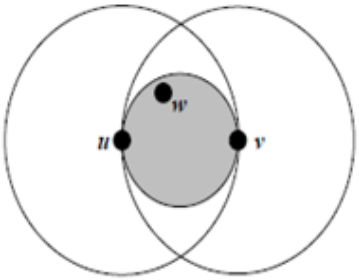


Figure 4: The GG graph. For edge (u,v)) to be included, theshaded circle must contain no witness w.

The Gabriel Graph is defined as an edge(u,v) exists between vertices u and v if no other vertex is present within the circle whose diameter is uv.

D. Protocol Implementation

We have made the following choices in our implementation to make GPSR robust on a mobile IEEE 802.11 network.

- **Support for MAC-layer failure feedback**: When a packet exceeds its maximum number of retransmit retries we receive a notification from

the 802.11 MAC layer. A retransmit retry exceeded failure indicates the recipient has left radio range. This feedback may inform GPSR earlier rather than expiration of neighbour timeout interval.

- **Interface queue traversal:** while a packet is retransmitted repeatedly at the head of its queue by an IEEE 802.11, it head-of-line blocks waiting for an acknowledgement from the receiver. The available transmit duty cycle of the interface reduces by this head-of-line blocking. We traverse the queue of packets for the interface and remove all packets addressed to the failed recipient upon notification of a MAC retransmit retry failure. We send these packets back to the routing protocol for re-forwarding to next hop. this eliminates MAC contention in high mobility simulations where neighbors were lost frequently.
- **Promiscuous use of the network interface:** In order to receive copies of all packets for all stations within its radio range the GPSR disables MAC address filtering. To reduce the rate at which packets must be sent and to keep positions in neighbor lists maximally as all packets carry their local sender's position.
- **Planarization of the graph:** We have implemented RNG and GG planarizations, through only RNG results are present in this paper. Both depend on having current position information for a node's current set of neighbors. In our implementation, every time a new neighbour is acquired or for a loss of a former neighbor we re-planarize the graph as evident by the data packet received from unknown neighbor or MAC transmit failure indication. In future, we try to keep the planarized graph maximally up-to-date.

III. PERFORMANCE EVALUATION

We have performed simulation using NS2(Network Simulator) to evaluate the performance of Greedy Perimeter Stateless Routing protocol. We have evaluated network metrics such as Packet Delivery Ratio(PDR), Routing Protocol Overhead(RPO) and Path Length. We also simulate Johnson stal's Dynamic source routing to compare the performance of GPSR.

A. Simulation environment

We set some parameters that constitute the context of our simulation in order to efficiently conduct different simulations. These parameters are defined in the following table:

TABLE II. SIMULATION PARAMETERS

TCL Configuration Variable	MobiCom 2000 Value
bint_	{1.0, 1.5, 3.0}
bdesync_	0.5
bexp_	3 * (bint_ + bdesync_ * bint_)
use_implicit_beacon_	1
use_mac_	1
use_peri_	1
use_planar_	1
use_timed_plnrz_	0

B. Simulation results

1. Packet Delivery Ratio (PDR)

Figure 5 shows for varying values of B, the beaconing interval, as a pause time, how many application packets can be successfully delivered by GPSR. In same figure, GPSR is compared with DSR. In graph, all algorithms deliver over 97% of source packets.

Only packets with destination address are included in the graph. A routing algorithm is not seen as failure if it tries to deliver to a truly disconnected destination. Since connectivity is dense, disconnection of a node is rare in these simulations. We get a slightly reduced delivery success rate for longer beacon interval of 3 seconds. GPSR delivers slightly greater packets than DSR when simulated in pause times.

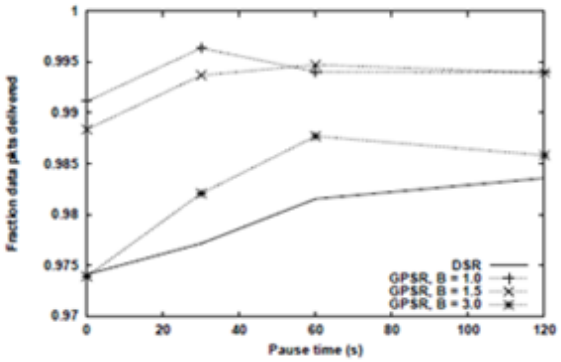


Figure 5: Packet Delivery Success Rate.

GPSR with varyingbeacon intervals, B, compared with DSR. 50 nodes

2. Routing Protocol Overhead

Figure 6 shows the Routing Protocol Traffic Overhead, during the entire simulation, for GPSR with varying B and for DSR the traffic Overhead is measured in total number of routing protocol packets sent through the network. For GPSR, even in different beacon interval, the results is a constant level of routing protocol traffic. Since GPSR's beacons are sent pro-actively. While DSR generates an increased routing protocol traffic as mobility increases since DSR is a reactive routing protocol.

In fig.6, we include a second DSR curve, DSR-Broch to simulate, our results for both GPSR and DSR shows 6 simulation runs.

In fig.7, GPSR with beacon interval b=1.5, gives between a threefold and fourfold overhead reduction under DSR. The graph shows that GPSR may give better savings in routing protocol overhead as mobility increases.

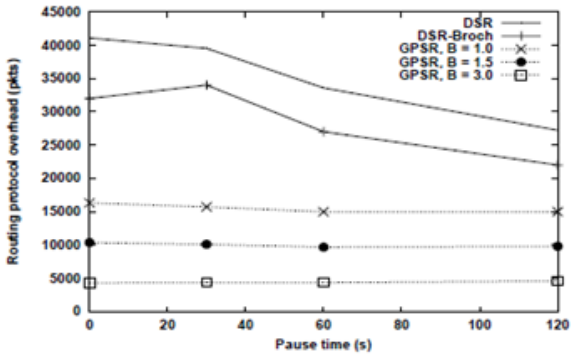


Figure 6: Routing Protocol Overhead.

Total routing protocol packets sent network-wide during the simulation for GPSR with varying beacon intervals, *B*, compared with DSR. 50

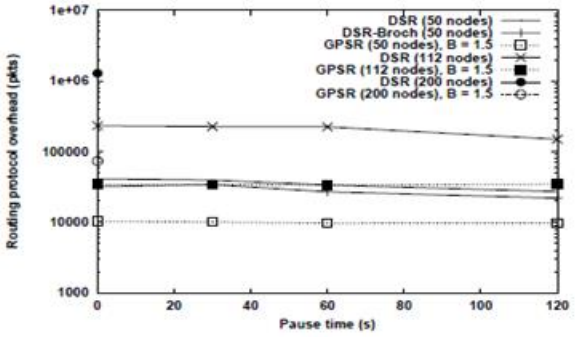


Figure 7: Routing Protocol Overhead.

Total routing protocol packets sent network-wide during the simulation for GPSR with B= 1:5 compared with DSR.yaxis log-scaled. 50, 112, and 200 nodes.

3. Path length

Figure 8,gives a histogram of the number of hop count over shortest path and the successively delivered packets by the GPSR and DSR. The data is shown as percentages of all packets delivered across all 50 node simulations of GPSR with beacon interval b=1.5 and DSR at pause time zero.

GPSR in optimal number of hops delivers the majority of packets. As shown in fig: Gpsr delivers

97% of its packets vs. 84% for DSR. This difference is due to DSR's caching which reduces route requests.

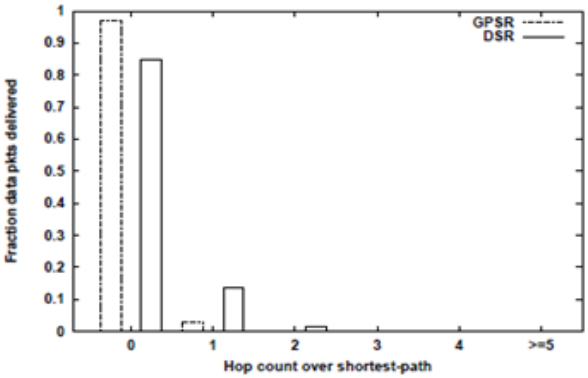


Figure 8: Path length beyond optimal for GPSR's and DSR's

IV. CONCLUSION

In this paper we have evaluated the performance of the GPSR protocol and have achieved good performance in following metrics: Packet Delivery Success Rate and Routing Protocol Overhead. Our simulations have shown that GPSR delivers upwards of 90% of data packets successfully and GPSR is very better than DSR as in high density of nodes as shown on 112- node and 200-node networks.

GPSR generates a constant less volume of routing protocol messages as nodes increases since GPSR generates routing protocol traffic independent of length of the routes in the network while DSR must find longer routes as distance of network increases and must often do as nodes increases. GPSR use of immediate neighbor information helps in end to end state and scaling other algorithms.

V. REFERENCES

[1]. Brad Karp, H. T. Kung, « GPSR: Greedy Perimeter Stateless Routing for Wireless Networks », ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 2000), 2000.

[2]. Brad Nelson Karp, "Geographic Routing for Wireless Networks", thesis in computer science, Harvard University, october 2000.

[3]. Mohamed AdneneZayene, Nabil Tabbane, Refaatelidoudi, "Performance evaluation of Greedy Perimeter Stateless Routing for ad-hoc networks , 2009 fourth international conference on computer science and convergence information technology, 2009.

[4]. Amina bengag, Asmebengag, Mohamed EL boukhari, "Enhancing GPSR Routing Protocol based on Velocity and density for real time scenario, 2020.

[5]. Yasser AL-roqi, Stylianos Papanastasiou, EvtimPeytchev, " Ferry assisted Greedy perimeter stateless routing protocol for mobile ad-hoc networks", 2014.

[6]. zhu Jing, Song binbin, Fu wei, Wang chengmin, "GPSR Protocol perimeter forwarding optimization algorithm based on Game model in UAV network", 2021.

[7]. Yanzhong Gao, jielin fu, yilu, "Improvement of GPSR routing protocol for TDMA- based UAV Ad-Hoc networks",2021.

[8]. A.Rodrigues, A.B Reis, S.sargento, "GPSR-PPU: Greedy perimeter stateless routing with position prediction and uncertainty for fanets", 2020.

[9]. Bashar Barmada, NajiAlobaidi, Iman ardekani, Guillermo Ramiezprado,MaryamE.Sabaee, "Evaluating vanet routing protocols for Auckland area", 2018.

[10]. Adam Walker, Milena Radenkovic, "GPSR-TARS Congestion aware geographically targeted Remote Surveillance for Vanets",2017.

[11]. Meriemhoumer ,MoulaylahcenHasnaoui, “An Enhancement of Greedy Perimeter stateless routing protocol in Vanets”, 2019.
[12]. Zhao li, bo yin, Mingling Gao, Xin Zhao, “Improved GPSR routing protocol based on weight function in shipborne ad hoc networks”,2020.
[13]. hosamrowaihy, Ahmed binsahaq, “Performance of GPSR and AOMDV in WSNs with uncontrolled mobility”,2016.
[14]. Chen chen, Huan li, Xiang li, Jianlong Zhang, Hong Wei, HAO Wang, “A geographic routing protocol based on trunk line in vanets”, 2021.
[15]. OhoudAlzamzami, Imad Mahgoub, “geographic routing enhancement for urban vanets using link dynamic behaviour:A cross layer approach”, 2021.
[16]. Ramin Karimi, Seed shokrollahi, “PGRP: predictive routing protocol for vanets”, 2018.
[17]. Linqi li, Xiaoyin Wang, Xinhua Ma, “design of a location-based opportunistic geographic routing protocol”, 2022.
[18]. ImaneZaimi, zineb squali houssaini, abdelaliboushaba, mohammedoumsis, “A new improved GPSR(GPSR-Kp) routing protocol for multimedia communication over vehicular ad hoc network”, 2016.
[19]. M.Ye, L.Guan, M.Quddus, “TDMP: Reliable target driven and Mobility prediction based routing protocol in complex vehicular Ad-Hoc network”, 2021.
[20]. Akashaharshad,Zurimamohdhanapi, shamala k. subramanium, rohayalatip, “Performance evaluation of the geographic routing protocols scability”,2019.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Study and Application of Decentralized System and AI in Metaverse

Ravi Kumar P, Ranjan Kumar Sah, Rehan Nasim, Supreme Pokhrel
Department of Computer Science & Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Due to the advancement in technology the world of virtual reality has become a reality. As COVID-19 spreads around the world, the need for the virtual reality has increased, and the metaverse industry is developing drastically. In the Metaverse, a virtual world that transcends reality, artificial intelligence and decentralized system are being combined. This chapter explains how artificial intelligence and decentralized system can affect the Metaverse.

I. INTRODUCTION

The word "Metaverse" is a combination of 'meta' which means 'virtual', and 'verse' which is a backformation from 'universe'. The Acceleration Studies Foundation, a non-profit research organization, categorized the Metaverse into four categories: a virtual world that experiences a flawless virtual story, a mirror world that reflects the current real world, an augmented reality that shows a mixture of augmented information in the real world and life logging, that captures and stores everyday information about people and thing. As the development of technology increases, the number of users who use the Metaverse increases, and due to activities at the same level as reality are performed, various kinds of large data are being generated. Data generated in the metaverse has specific value for themselves. In the Metaverse, the amount of data is directly proportional to the value, and the importance of reliability and security is increased because of this. Decentralized system is required to guarantee the reliability of data in the Metaverse, and artificial intelligence is used to secure the diversity and rich content of the Metaverse. The contents will be developed in the following order.

In this chapter, under the theme of the Metaverse, we will investigate the issues of human instinct for creation in the virtual world, the phenomenon in which the real and the virtual are combined in the virtual world, and the reliability of data in this virtual world. Blockchain and NFT technologies are described as trust technologies. And the Metaverse platform built based on this technology will be described. We will understand the interface between blockchain and artificial intelligence and look at how a better world is created by combining blockchain and artificial intelligence in Metaverse.

II. VIRTUAL WORLD AND DESIRE OF CREATION

A. Human desire for creation

Humans have an instinct for creation, and this creativity is a crucial factor that distinguishes humans from other animals. The creativity of human beings has been creating the culture. The paper published in 2004 described the SeaCircle as the new concept of the culture, and it regarded the SeaCircle as human cultural activities for creating. In the concept of the SeaCircle, humans are the spiritual beings, and only humans constitute a culture. It explains the elements of insight of culture. According to the SeaCircle theory, creativity is explained as an element of Open Mind and Spirit.

On the SeaCircle concept, the Metaverse can be interpreted as a space that allows people to be more immersed in creative activities by resolving some of the constraints on space and resources.

B. Link between the real and the virtual world

In current years, the digital and bodily worlds have turned out to be increasingly intertwined. The First and Second Industrial Revolutions has been maximising performance via labour division, which supposed that fabric manufacturing and intake had been separated. As the Third Industrial Revolution progresses, records have turn out to be a precious commodity, and offline transactions are step by step being supplanted through on-line transactions. As matters and human beings turn out to be hyper-related during the Fourth Industrial Revolution, a smart revolution is taking place. Such as social customisation or virtual DIY, there may be a convergence phenomenon wherein manufacturing and intake arise on the identical time (Design It Yourself).

Due to restrained resources, the offline global of substances is ruled through Pareto's law, which strives to personal and deal with the centre of 20%. The Long Tail idea, on the alternative hand, is used withinside the on-line realm of facts to percentage and locate possibilities from the marginalised eighty percentage of clients. The Fourth Industrial Revolution is bringing offline and on-line worlds together. Manufacturing, logistics, finance, automotive, sports, healthcare, education, cuisine, and normal existence are all experiencing this convergence.

C. The concept of real and virtual in Metaverse

The Metaverse, which is modelled after the real world, contains political, economic, social, and cultural connections. The process of interworking and convergence between the physical world and the Metaverse is depicted in Figure 1. The Metaverse represents another reality that is impossible to accomplish in the real world.

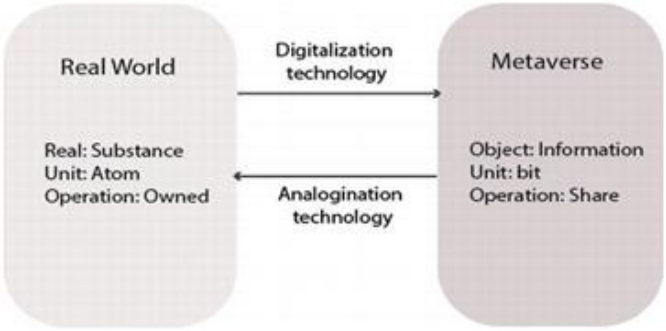


Figure 1: Relationship between the real world and the Metaverse.

Due to the difficulty of going to school due to COVID-19, UC Berkeley students established a campus inside the Minecraft game and hosted an event to perform a virtual graduation ceremony. Even the ritual of throwing hats after graduation was re-enacted in Minecraft, with the president, guest speakers, and graduates all dressed up as Minecraft characters.

Instead of going to an offline office, game developers can build games on the Roblox virtual platform. Tokens are collected in Roblox in exchange for

labour, and the tokens earned in the game can be taken outside to be cashed.

Thousands of users can freely trade products and services in Metaverse using the platform's currency and transaction method. Both The Sandbox's virtual asset SAND and Digital Land's virtual asset MANA are listed on the exchange and regularly traded. This means that virtual money in the form of bits can be used to substitute money in the actual world. Co-creating a game in the virtual world requires the virtual world's space to be replaced with the real world's space. This means that real-world activities are represented in virtual space as data in bits, with the importance and trustworthiness of this data highlighted.

III. DATA TRUST

The virtual world has evolved rapidly since the fourth industrial revolution. The real thing has been translated into data from the virtual world, and the virtual world has even influenced the real world's direction. We have a question here about data veracity, and if the real thing is appropriately becoming data in the virtual world. Trust technology is becoming increasingly crucial in the virtual world. Blockchain is an example of one of these trust technologies.

Satoshi Nakamoto first introduced blockchain in his paper "Bitcoin: P2P Electronic Money System" in 2008. Blockchain can be defined as a technology that ensures the security of individual transactions. A blockchain is made up of data-containing blocks connected by a chain. To construct and connect blocks, a blockchain is employed, and consensus methods are used in the process. Any node in the network can build blocks, but not all generated blocks are connected, and only one block is recognised and connected. Because only one block out of several is connected to the previous block, and the others are discarded, consensus among participating nodes is required to choose one block. Consensus algorithms such as Proof of Work (PoW) and Proof of Stake (PoS) are used to obtain a consensus. If all nodes recognise it as a unique block, the node who created it will be rewarded with cryptocurrency. Mining is the term for this operation, and a Canonical Chain is a blockchain that is exclusively connected by blocks created by mining.

Blockchain is emerging and evolving. With the introduction of Bitcoin, Blockchain 1.0 was a time of financial system revolution. Bitcoin is significant because it aims to create a unified global financialsystem based on decentralisation, which is basic blockchain concept. Ethereum smart contracts are at the centre of Blockchain 2.0, which is a time of contract automation. It made it feasible to execute legal contracts online using only computer code and no transaction intermediary. It's a time when the platform's potential for growth as an online trading platform was demonstrated. Blockchain 3.0 refers to the point at which blockchain technology has been widely adopted and implemented across a variety of industries. Technological enhancements, including as changes in consensus algorithms, faster transaction processing, and in-house decision-making mechanisms, are being implemented to alleviate the shortcomings of the previous blockchain. While artificial intelligence is likely to be applied to a wider range of areas in Blockchain 3.0, more diverse blockchain and AI applications are predicted to emerge in the Metaverse environment.

Figure 2: Blockchain paradigm evolution direction.

IV. METAVERSE BASED ON DECENTRALIZED SYSTEM

A. Ethereum code

Ethereum is a platform network that uses its own blockchain to run a variety of decentralised apps (DApps). Ethereum Request for Comment (ERC) documents the intricacies of Ethereum in the same way that RFCs document the basic structure and features of Internet standards. The basic procedure for issuing tokens in DApps using the Ethereum network is stated as an ERC-number, as illustrated in Table 2. When DApps using the Ethereum network issue tokens, they must follow the ERC-number protocol. ERC-20, ERC-165, ERC-223, ERC-621, ERC-721, ERC-777, ERC-827, ERC-884, ERC-998, ERC-1155, ERC-1404, and so on are all Ethereum standard papers.

ERC-20 is a protocol for replaceable tokens, and ERC-20 tokens have the same value and function as each other and may be swapped. The Ethereum project creates ERC-20 tokens that can be used for investment and a variety of companies.

ERC-721 is an NFT protocol (Non-Fungible Tokens). NFT ensures uniqueness by permanently storing encrypted transaction history on the blockchain. Each token has a unique recognition value that authenticates digital asset ownership and assigns a value to the transaction. NFT has primarily been used to commemorate unique moments or collect digital assets, but it has recently merged with Metaverse to form a new digital content business.

ERC-20	ERC-165	ERC-233
ERC-621	ERC-721	ERC-777
ERC-827	ERC-884	ERC-1155

Figure 3: ERC (Ethereum request for comment)-number.

B. Metaverse based on decentralized system

The Metaverse is a three-dimensional virtual arena where people engage in social and economic activities in the same way they do in real life. Within the Metaverse, NFT serves as a medium for interaction and proof of private property. CryptoKitties is an example of an NFT application. It is a cat reproduction game based on the blockchain. CryptoKitties is a DApp built on the Ethereum ERC-721 token. In CryptoKitties, each player is given only one cat in the globe. Cat digital assets are special because, unlike other virtual assets, they have a separate unique recognition value. When a service is cancelled in general online games, there is a difficulty in that the character formed in the game can no longer be owned. Individual members linked to the network, on the other hand, can distribute and store digital assets using NFT technology to establish ownership.

Decentraland used virtual reality and blockchain technologies to bring the concept of real estate to the Metaverse. Using MANA, an ERC-20 token, Decentraland made it possible to buy land, or virtual real estate. Users can freely construct structures on Decentraland-purchased land, earn money by attaching billboards to structures, and host exhibitions by collecting rare digital information. ERC-721 non-fungible tokens are used to represent land ownership and other collectible items. These one-of-a-kind assets are created using Ethereum smart contracts, which enable owners to prove ownership on the blockchain ledger. MANA is a cryptocurrency that can be bought on exchanges and used to buy digital products and services all over the world.

Enjin Coin is a cryptocurrency project that was designed for the purpose of trading game items and is also an integrated platform for developing blockchain-based games. Enjin is a smart contract platform based on the Ethereum blockchain, as well

as a protocol and cryptocurrency that allows game developers, content creators, and gaming communities to create, manage, and implement virtual goods. A mobile cryptocurrency wallet can store and handle Ethereum-based ERC-20, ERC-721, and ERC-1155 token objects. The ownership and financial worth of game items utilised in all games are guaranteed by Enjin Coin. When Enjincoin is utilised as a currency in the Metaverse, it can be used not just in the Metaverse, but also in the real world with currency value.

V. COMMON ASPECTS OF DECENTRALIZED SYSTEM AND ARTIFICIAL INTELLIGENCE

A. Artificial intelligence and decentralized system

Humans have progressed to the stage of connected intelligence, which employs the united intelligence of humans and machines, thanks to the cognitive, agricultural, and scientific revolutions. A symbiotic relationship between humans and machines has begun, much as in the movie The Matrix, and artificial intelligence and decentralized system technology are driving it.

Artificial intelligence has progressed to the point where massive volumes of data and pattern recognition can be used to forecast and create. Artificial intelligence is also assisting individuals in reducing the number of repetitive jobs and human errors. Decentralized system technology has infiltrated our culture as a digital asset, and via decentralisation, it is evolving into a secure and dependable transaction. Artificial intelligence (AI) is at the heart of the Fourth Industrial Revolution, and it may be used with decentralized system technology to increase the power of both AI and blockchain. Artificial intelligence and blockchain have the potential to disrupt business models and transform society.

B. Decentralized system for artificial intelligence

Because artificial intelligence is centralised and data is managed and stored centrally, it is vulnerable to hacking and manipulation, which can lead to data tampering. Furthermore, there are several flaws and hazards because the source and trustworthiness of the source for creating data are not assured. The immutability, origin, and control mechanisms of blockchain have the potential to overcome artificial intelligence's flaws and improve trust, privacy, and decision-making accountability. The combination of decentralized system and artificial intelligence can aid in trusted data analysis and decision-making. It can also be used to provide safe data sharing, explain artificial intelligence, and regulate trust amongst devices that can't trust each other.

C. Artificial intelligence for decentralized system

The data on the blockchain is guaranteed to be accurate. The security of apps built on top of the blockchain technology, on the other hand, is questionable. Furthermore, when a new block is added to the blockchain and all nodes must agree, an issue occurs in that which makes it inefficient for industries that need fast speed. When an error or vulnerability in a smart contract's script is discovered and needs to be fixed, the blockchain's irreversibility can make this difficult. The recent hacking of tens of millions of dollars in crypto currencies utilising flaws in smart contract algorithms highlights the need for agents that can adjust for algorithms that are not flawless right away. Artificial intelligence machine learning systems can increase the security of blockchain applications, alter dynamic parameters for scalability, and provide effective personalization and governance mechanisms in such circumstances.

Netflix offers a list of related films to your favourite films; however, this is the result of Netflix's central server analysing personal data. Your personal

information will be protected if you do not disclose it to Netflix, but you will not be supplied with a personal taste analysis. Rather than gathering data on a single server, you might use data saved on a decentralised blockchain. However, because anybody may look at the transaction record on a public blockchain, there is a risk of privacy invasion as well. Although allowing individuals to directly handle personal information on the blockchain is viable, there is a risk of incurring significant costs. Individualized services can be provided by artificial intelligence without compromising personal information. Artificial intelligence can conduct analysis on the user's local device while avoiding any analysis that has not been approved in advance. Artificial intelligence can enable decentralisation, giving people real control over their personal data.

VI. DECENTRALIZE SYSTEM AND ARTIFICIAL INTELLIGENCE IN THE METAVERSE

Decentralized system is critical to the Metaverse's economic system's implementation. Without blockchain, someone will eventually dominate the Metaverse economy. If the decentralized system is not supported, it will be difficult for resources or items utilised in the Metaverse to be valued or to have economic transactions that are comparable to those in the real world. The Metaverse was further activated using NFT-based decentralized system technology. Metaverse becomes the world to realise it with the introduction of WEB 3.0 and Blockchain 3.0.

People emerge in the Metaverse by scanning themselves in three dimensions or morphing into avatar figures. Characters in the Metaverse, like clones in real life, are acknowledged as more than just game characters. Aside from their own avatars, users construct objects in the Metaverse that can represent their individuality. The blockchain's NFT technology is utilised to demonstrate this.

A. High quality learning data

Artificial intelligence in Metaverse may readily replace the problem of people's time, work, and expense in the actual world. When it comes to providing news in the real world, you must put in a lot of effort, such as finding an announcer, shooting in a studio, and editing footage. However, in the virtual world, an artificial intelligence announcer can convey vital and significant news swiftly and constantly for an extended period. It is vital to master the facial expressions, muscle movements, voices, accents, and gestures of genuine announcers to convey news in the Metaverse. We extract only the part where the voice of another reporter and noise-free data, and the announcer's face and voice come out clearly toward the camera, when learning by receiving a long-time news video from a broadcaster to make an artificial intelligence announcer video, and only detect a specific person technique must be applied when learning by receiving a long-time news video from a broadcaster to make an artificial intelligence announcer video. When looking for varied data, such as this, you can leverage blockchain meta-information to choose only the pure data required for learning and induce high-quality learning. The essential high-quality data is selectively available thanks to metadata recorded within the blockchain block. It is created in the Metaverse as dependable data, increasing the number of consumers who use the Metaverse.

B. Reusable data

Artificial intelligence, rather than human people, has recently been used to produce creative activities in the Metaverse. Artificial intelligence artists study about the trends and styles of the works before expressing what they've learnt in their work. For style

analysis in the past, a large amount of data was used. Now, data is stored in a distributed ledger so that it can be easily selected and reused by artificial intelligence artists. Acquiring more data and practising iteratively minimises the likelihood of selecting incorrect data and speeds up validation.

C. Stable decentralized network

Metaverse is a virtual three-dimensional world that necessitates a lot of data and server resources. Controlling through a central server, on the other hand, can be expensive. It is required to have a network system that can use the Metaverse environment with everyone's PC computing while using the distributed environment system of blockchain. The burden of centrally handling enormous amounts of data can be lessened when users have control over the Metaverse environment they want to utilise or view. It also has the potential to prevent some large tech businesses from monopolising the Metaverse.

D. Privacy

There is a need for a system that can govern in such a way that ethical issues with Metaverse residents do not develop. Only publicly available data on real and virtual people should be made public. To protect privacy, a personal information security system should be implemented. However, because digital virtual individuals have no legal standing, they are easier to manipulate or modify than real people, and there is a risk that the wrong algorithm will be used, resulting in major racial and gender discrimination. In terms of personal information, blockchain may be used to safely protect it from external threats. If personal information is incorrectly altered, it can be appropriately managed by following a clear path that can be tracked back to the time of occurrence

E. Distinguishing between virtual and real

A gadget that can distinguish between artificial intelligence and actual people is required to create a stable Metaverse environment in which users are not confused. From a human perspective, the fictional characters utilised in the Metaverse have progressed to the point where it is difficult to tell the real from the false. To guide the comparison and evaluation of real and fictitious persons, a credible data creation system is required. So that phoney news and fake photos can be discovered, data should be transparent and informative. Data content should be maintained in a blockchain so that individuals can accurately know and interpret artificial intelligence-generated data, as well as the detailed history if needed. Decentralized system technology can be utilised as a data source to explain artificial intelligence-generated data.

F. Rich content

Artificial intelligence technology is being used to mimic and replace human behaviour. Artificial intelligence analyses user behaviour patterns in the Metaverse, such as words and messages, to forecast a user's personality, intellectual level, and economic level. Metaverse creates human-like voices and original material using artificial intelligence. By simply inputting simple information, these data can be instantly turned into games, YouTube, news, ads, and instructional materials. Using artificial intelligence technology and the huge data required for the Metaverse universe, it is possible to construct vast pattern content that imitates human behaviour. Personal information may be safely stored on blockchain, and other sorts of material can be developed more easily.

G. Economic virtuous cycle

Artificial intelligence can be used in investment and business to make decisions about which data to employ. In updating forecasts, it is critical to have more dependable data. When blockchain data is employed, more reliability can be ensured through historical management, which improves the accuracy of business predictions. In addition, the Metaverse Marketplace can be further activated by using blockchain technology to pay for tokens and coins

VII. CONCLUSION

In the Metaverse, various and copious amounts of secondary and tertiary data are generated due to the activities of many users. In the decentralized system based Metaverse, this data has a unique identification tag and is used as traceable data. Such data is becoming a good material for artificial intelligence in the Metaverse. Metaverse uses artificial intelligence and decentralized system technology to create a digital virtual world where you can safely and freely engage in social and economic activities that transcend the limits of the real world, and the application of these latest technologies will be accelerated. Artificial intelligence and decentralized system technology are expected to play an essential role in the ever-expanding world of the Metaverse.

VIII. REFERENCES

[1]. J. M. Smart, J. Cascio, and J. Paffendorf, Metaverse Roadmap Overview; Acceleration Studies Foundation; 2007

[2]. Oh Jeong-seok, Youn Ho-chang, A Study on New Concept of Culture; Proceedings of the Korea Contents Association Conference; 2004. p.54-60

[3]. Oh Jeong-seok, Youn Ho-chang, Jean Hyun-joo, Kim Tea-heon, A Study on a Creativity of SeaCircle; Proceedings of the Korea Contents Association Conference; 2013. p.61-62

[4]. Lee Min-hwa, Examples of Advanced Countries in the Forth Industrial Revolution and Korea's Response Strategy; Advanced Policy Series; Hansun Foundation; 2017. p.14-107

[5]. Gretchen Kell, Unforgotten: COVID-19 era grads to be celebrated virtually this Saturday, Media Relations, Berkely News [Internet]. 2020. Available from: https://news.berkeley.edu/2020/05/14/unforgotten-covid-19-era-grads-to-be-celebrated-virtually-this-saturday [Accessed:2021-01-30]

[6]. Roblox Studio, anything you can imagine, make it now! Roblox [Internet]. 2021. Available from: https://www.roblox.com/create [Accessed:2021-01-30]

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Modelling of Unmanned Ground Vehicle Using Neat

Rajesh G[1], Saroja B[2], S Ramalakshmi Malladi[3]

[1]Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru 560103, Karnataka, India

[2]Department of Electronics and Communication Engineering, Siddartha Institute of Science and Technology, Puttur 517581, Karnataka, India

[3]Department of Mechanical Engineering, Raghu Institute of Technology, Vizag 531162, Andhra Pradesh, India

ABSTRACT

Self-propelled automobiles are not merely futuristic possibilities; they are an integral element of our lives. The purpose of this article is to present a better lane detection technique based on the neuro evolution of augmenting topologies. The purpose of this work is to provide a robust end-to-end lane recognition approach for Unmanned ground vehicles using neural network techniques. By combining edge detection and polynomial regression, a baseline strategy for recognising just straight lane lines is achieved. This latter section detects both straight and curved lane lines. The proposed lane detection approach outperforms the standard strategy in the NEAT scenario.

Keywords: Unmanned Ground Vehicle; Edge detection; Polynomial Regression; NEAT;

I. INTRODUCTION

Self-driving vehicles are becoming a major industry in almost every country. After WWII, the global automobile stock topped 80 million, and then 90 million by 1960. It was 130 million after five years, 291 million in 1980, 419 million in 1990, and 731 million in 2011. According to projections, it will reach two billion by 2020. For hundreds of years, this sector has been fueled by steam, then petroleum, and diesel, and now it appears that electric driving will be the future. Obviously, faster and more valuable vehicles can be produced, but in our fast-paced world with ever-increasing vehicle numbers, the frequency of accidents has increased. As a rule, the driver is to blame for these catastrophes, which might theoretically be remedied by Unmanned ground vehicles. Currently, human presence is the most important factor in transportation, despite the fact that there are several places where a device or feature might help people perform better. Some examples are the autopilot in an aeroplane, the cruise control in a car, and various other devices that help dynamics. In this test, we will examine the development of unmanned ground or driver assistance devices and how people feel about them.

1. Unmanned Ground Vehicles

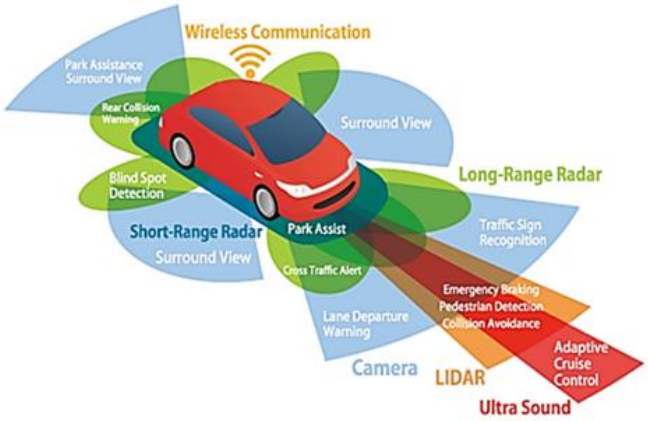
The National Highway Traffic Safety Administration of the United States categorises unmanned ground vehicles into six levels of autonomy. Humans are in

charge of the driving at Level 0, which progresses through driver aid technology to completely autonomous vehicles. As illustrated in figure 1, the vehicle's ADS acts as a virtual chauffeur, taking complete control of the vehicle and driving it in all conditions.

Instead of autonomous, the SAE [Society of Automotive Engineers] uses the term "automated." A human passenger must always be present in order for a self-driving car to function properly. Level 3 (conditional driving automation) or Level 4 unmanned ground vehicles would be appropriate (high driving automation). There are two types of vehicles in Figure 2, the Unmanned ground vehicles (also known as driverless cars or autonomous cars) and the Unmanned ground vehicles (also known as autonomous cars or driverless cars) (AI). Autonomous vehicles can only operate on roads that haven't been modified for their use in order to be called truly autonomous.

AI technologies power unmanned ground vehicle systems. Engineers employ a combination of image recognition, machine learning, and neural networks to handle massive amounts of data in order to develop unmanned ground vehicles. The neural networks provide the data into the machine learning algorithms, which uncover patterns in the data. As a result, a neural network may learn to detect traffic signals, people, street signs, and other components of a given driving environment by analysing photographs acquired by cameras mounted on unmanned ground vehicles.

Google's Waymo project is an excellent example of a fully autonomous unmanned ground vehicle. However, the human driver is only needed to override the system when necessary. Even if a vehicle's autonomy is limited, self-driving technology can be found in many of the consumer models available today.



The issues vary from technological to legislative to philosophical. State-by-state autonomous vehicle rules could lead to "zombie cars" on the road. Unmanned cars might reduce urban CO2 emissions by 80% by 2050. Autonomous vehicles might be fully realised if three trends were implemented concurrently: Ridesharing, automation, and electrification.

Sensors used by Unmanned ground automobiles include radar, lidar, sonar, GPS, audiometry, and inertial measurement units. However, none of them are as efficient as the genetic algorithm NEAT because it does not use live data. Supervised machine learning is a subset of machine learning and artificial intelligence. Despite the substantial variance in the prior, the genetic algorithm surpasses reinforcement learning on mean learning time. Genetic algorithms are used in artificial intelligence, just as other search algorithms, to search a space of potential solutions for a problem.

II. LITERATURE REVIEW

PéterSzikora et. al. [1] proposed that Self-propelled cars are not just opportunities of the future, but are an important part of our present. The article describes the steps and types of self-driving cars and their impact on people. It also looks at the human side of this rapidly developing, therefore very exciting topic.

Venkata Shiva Prasad Nannuri et al. [5] showed a robust end-to-end lane detection approach on the Raspberry Pi using Python and Opencv, They proposed a straightforward method based on edge detection and polynomial regression. Then, using viewpoint modifications and histogram analysis, we present an improved lane detecting technique.

Mochamad Vicky Ghani Aziz et al. [2] described the results of a lane detecting algorithm deployment on the Cipularang toll road. The image processing approach is a combination of colour region, line selection, canny edge detection, and Hough transform methods. Overall, the Python implementation approach can successfully recognise the road lane with an accuracy of more than 90%.

Mahmoud Fathyaet al. [3] highlighted how self-driving vehicles are nearing legalization. They are, however, not trustworthy enough to be employed in real life. A self-driving automobile prototype that integrates several technologies is suggested. A unique road setting was developed to fit the automobile prototype in order to test it.

Wael Farag et al. [4] have developed and implemented a highly accurate method for detecting and tracking traffic lanes. Self-driving cars or Advanced Driving Assistance Systems applications can benefit from this technology. The primary focus of the proposed method is to identify driving lanes with high precision and predictability.

III. PROPOSED METHODOLOGY

This work advises the simulation of Unmanned Ground Vehicle (UGV) in 2D using NEAT (Neuro Evolution of augmented Topologies). The simulation is instituted on UGV for collision avoidance with respect to road borders. The UGV's sensors collect the data and fed to neural network for training. The vehicle's performance is quantified by a fitness value. This will grow from group to group with a larger fitness value. The neural trained vehicle will follow a randomly generated path thereby avoiding road borders, and go as quickly as possible.

Neurons are coupled by connections with weights ranging from -1 to 1. These neurons are arranged into layers, the first being the input layer and the last being the output layer, with values ranging from 0 to 1. 8 of the 9 input neurons represent the car's sensors that measure distance from the road edges. They are 1 if the wall touches the car and 0 otherwise. The ninth neuron represents the car's velocity, ranging from 0 to its top speed. The information transmitted through the connection is the product of the node value and the connection weight. Receiving neurons utilise a specific function to combine input values into an output that is then forwarded across the network. These commands instruct the car to accelerate, brake, or turn left or right depending on the value of the output node.

For a network capable of driving a car, the hidden layers are filled with nodes, and connections are weighted appropriately. Python's NEAT library is used to do so. NEAT stands for Neuro Evolution of Augmented Topologies.

Starting lines build many primitive neural networks, each controlling an automobile. This is because these networks are produced at random. Every car gets a

"Fitness Value" score. Their fitness value increases if they avoid obstacles and stay on track longer than the others. After the run, the best networks are picked and allowed to generate children, who will be the next run's pilots; these vehicles will restart on the generated path and perform better than the previous generation. Repeat until the NEAT's augmented topologies portion is reached.

A child inherits traits from both parents as well as new traits, such as a different value in a connection here or a new neuron in the hidden layers. As new networks get more complex, their structure (topologies) expands to achieve the optimum fitness value. But it will take a long time to train. Change their top speed and turn radius to increase the time. In order to turn left or right, they will need to maintain steady acceleration. Bicycles, crossroads, traffic lights, etc. to spice it up are also added.

IV. IMPLEMENTATION AND RESULTS

Python is an interpreted high-level general-purpose programming language. Its syntax and object-oriented approach are intended to assist programmers in writing clear and logical code. PyCharm is a graphical integrated development environment (IDE) for computer programming, particularly for the Python programming language.

1. Evaluation Measures

The following are the statistical measures to evaluate the performance of a given classifier. The Network Parameters are neurons at input and output.

```
INPUT_NEURONS = 9
OUTPUT_NEURONS = 4
```

The Input and the output neurons: These neurons have weight between -1 and 1 higher the weight brighter the color of the neuron.

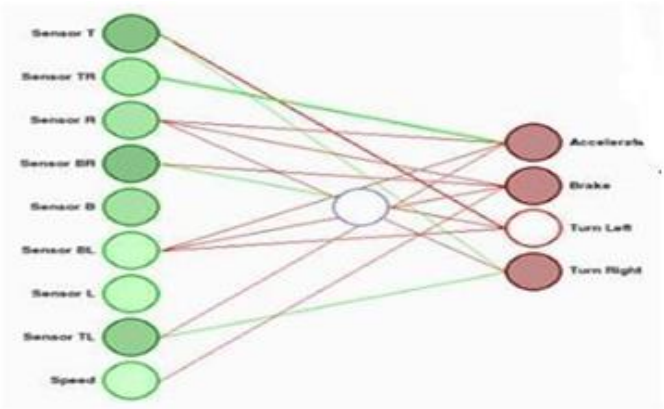


Fig 4. Implemented Neural Model

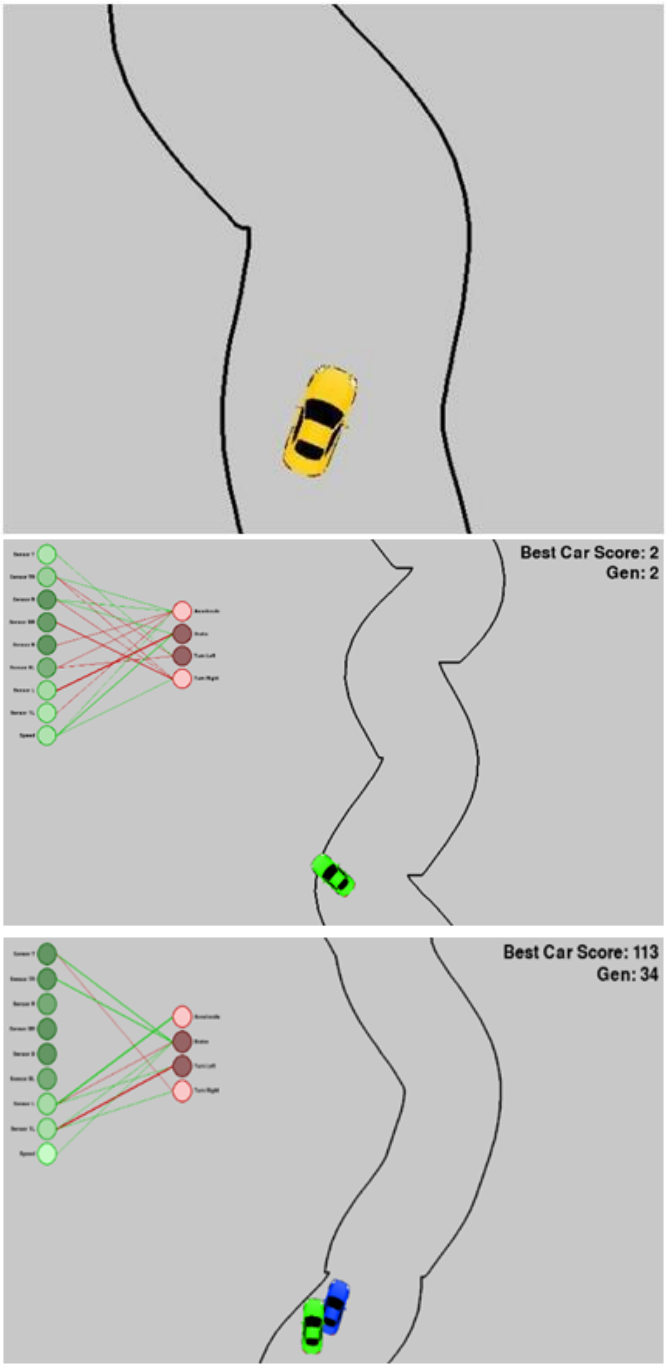
The population size and the threshold are modified under the NEAT section, figure 5. The simulation will end when any genome reaches the fitness value. It begins with one car in generation '1' and gradually evolves from generation to generation with a higher fitness value.

```
[NEAT]
fitness_criterion     = max
fitness_threshold     = 1000
pop_size              = 50
reset_on_extinction   = False
```

Fig 5. Fitness Generation

As seen in Figure 7, the car in generation 2 crashes extremely early and has a fitness value of only 2. This is due to the fact that the car in generation 2 has a low fitness value. Figure 3 depicts the generation 3 fitness value, in which the car achieves the highest possible score of 113 points. This method will be applied to real-world automobiles by selecting the network that has achieved a fitness value of more than 1000 and adjusting the value in accordance with the needs of the current world. Furthermore, because the real-world car has the same setup settings as the virtual car, this network will be extremely effective. Because of the slowed speed of the vehicle as well as the

increased turning radius, this algorithm becomes more effective.



V. CONCLUSION

This simulation-based system employs NEAT and is capable of recognizing straight lanes, curved lanes, and tight corners. This code should also be effective at avoiding obstacles because it has a brake neuron that causes it to stop when it senses something in front of it. Unlike supervised learning, the genetic algorithm uses live data rather than past data, and if there are any new changes in the road (lane), the supervised learning algorithm is ineffective because it works on datasets from the past, whereas the genetic algorithm adapts to those changes and acts accordingly. As a result, (NEAT) is one of the most successful algorithms for tackling traditional reinforcement learning problems, and it has proven to be quite effective in this article.

It is more challenging for reduced road width and higher number of turns or make the turns close-fitting. The more generations that are added to the network, the more sophisticated and complicated the hidden layers will become. Further research is being conducted for improved efficiency by adding obstacles such as walking humans, bicycles, and vehicles, and modifying the code to avoid obstacles by changing the direction or halting until the obstacles start moving away from the car; also, by introducing traffic signal rules and reading other traffic sign boards such as stop signs, speed breaker signs, different zone signs, and so on; and providing vehicle to vehicle communication for multiple cars for more data.

VI.REFERENCES

[1]. P. Szikora and N. Madarász, "Self-driving cars — The human side," 2017 IEEE 14th International Scientific Conference on Informatics, 2017, pp. 383-387, doi: 10.1109/INFORMATICS.2017.8327279.

[2]. M. V. G. Aziz, A. S. Prihatmanto and H. Hindersah, Implementation of lane detection algorithm for self-driving car on toll road cipularang using Python language, 2017 4th

International Conference on Electric Vehicular Technology (ICEVT), 2017, pp. 144-148, doi: 10.1109/ICEVT.2017.8323550.

[3]. Mahmoud Fathy, Nada Ashraf, Omar Ismail, Sarah Fouad, LobnaShaheen, Alaa Hamdy, Design and implementation of self-driving car, Procedia Computer Science, Volume 175, 2020, pp165-172, https://doi.org/10.1016/j.procs.2020.07.026

[4]. Mr.MustafaSurti, Bharati Chourasia, Real time lane detection system using Python and OpenCV on Raspberry Pi , International Journal for Research in Engineering Application & Management,Vol-05, Issue-06, Sep 2019.

[5]. Abdulhakam.AM.Assidiq, Othman O. Khalifa, Md. Rafiqul Islam, Sheroz Khan, Real Time Lane Detection for Autonomous Vehicles , Proceedings of the International Conference on Computer and Communication Engineering 2008 , May 13-15, 2008 Kuala Lumpur, Malaysia .

[6]. KorneliaLazányi, "Are we Ready for Self-Driving Cars-a Case of Principal-Agent Theory", Applied Computational Intelligence and Informatics (SACI) 2018 IEEE 12th International Symposium on, pp. 000251-000254, 2018.

[7]. Yunping Zheng, Ruijun Li, Wenjie Huang, Mudar Sarem, "A NAM-based Edge Detection Algorithm", Natural Computation Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2018 14th International Conference on, pp. 1154-1159, 2018.

[8]. Mohamed Azab, Bahaaeldin Mohamed A. Moustafa, "Multi-Mode Self-Driving EV Control for Industrial Sites and Logistic Stores", Novel Intelligent and Leading Emerging Sciences Conference (NILES) 2020 2nd, pp. 41-46, 2020.

[9]. D. Ramesh Reddy, ChandravathiChella, K. Bala Ravi Teja, Heera Rose Baby, Prakash Kodali, "Autonomous Vehicle Based on Deep Q-Learning and YOLOv3 with Data Augmentation", Communication Control and Information Sciences (ICCISc) 2021 International Conference on, vol. 1, pp. 1-7, 2021.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Embedded Web Server

Sanjana R Chandragiri, Om Shankar Sah
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, India

ABSTRACT

The embedded web server technology is the mixture of embedded device and internet era, which offers a flexible remote tool monitoring and control function based totally on internet browser and it has become an advanced development fashion of embedded generation. thru this embedded web server person can get entry to their equipment remotely. The equipment noted here will be domestic home equipment and manufacturing facility devices. This paper is targeted on awareness of TCP/IP suite and user improvement platform for this embedded net server. The embedded web server layout consists of a complete internet server with TCP/IP aid and Ethernet interface. it also includes guide for sending mail, and software for computerized configuration of the web server inside the network. The web server reference design consists of complete supply code written in C-language.

Keywords: domestic home equipment, computerized configuration.

I. INTRODUCTION

Of all the semiconductor industries, the Embedded systems marketplace is the most conservative. Engineering decisions in this market are usually conservative, leaning towards established, low-risk solutions. Because of this, the basic infrastructure of embedded systems has only evolved slowly over the past ten years.

Intelligent homes will be connected to the Internet and requires a microcontroller to communicate with the other network devices. Recently, ubiquitous computing technology and home network technology have developed rapidly and are in practical use. Therefore, more and more convertibility to the Internet will be required for home appliance such as lighting fixtures, refrigerators and audio-visual equipment. When these appliances are fitted with servers which are connectable to the WWW, interactive communication will become possible and new functions and services will become available. The embedded web server can simplify the design process for embedded web server applications. There has been exponential growth of Internet use in
recent years. This has generated a strong trend toward using Internet protocols.

II. USES

A. System Description

The embedded net server reference design is designed for integration in virtual equipment. The embedded internet server can be plugged into any Ethernet interface and talk with a popular net browser. figure 1 illustrates a few situations in which the net server may be used.

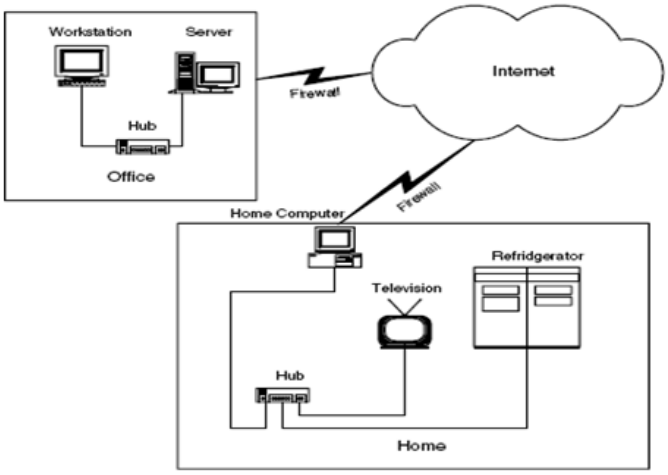


Fig.1 Monitoring Home Equipment from Office

Figure.1 explanation: As shown in discern 1, numerous patron electronics may be controlled from a pc connected to the internet. The net web page is the "manage center" for the embedded net server. think the embedded net server is embedded in several devices in a residence. every server is attached to the community. A pc placed at home as on determine controls all gadgets and might receive requests from other computers at the net. The internet server is recognized by way of its particular IP deal with and can be controlled remotely from anywhere within the global so long as the authorization is so as.

B. Hardware Structure

Excessive performance chip changed into produced abundantly each yr. The price is appropriate. The high attention is the size and gadget value. Many embedded computer systems are bodily placed within a few huge Arti-facts. The embedded internet server reference design is designed to be flexible for future improvement. The net server commonly communicates via an Ethernet connection. However, it is also capable of speak with SLIP and modem connection the use of the built–in UART. A CPLD is cover within the reference design to allow memory mapping of other gadgets to the system. With best SRAM and the Ethernet controller related to the statistics bus the CPLD can be not noted. The hardware design is flexible, making it usable for plenty applications. it is designed for use with Ethernet, SLIP or PPP connections, either through a LAN, any other pc or through a dial-up connection. it is also viable to attach other external peripherals to the machine. The on- board Flash reminiscence can be expanded without hardware modifications.

C. Memory

The web server includes enough memory to develop large applications on top of the web server protocols. 32K bytes of external SRAM is used for buffering data. A 2-Mbit External Data Flash is used for storing web pages to allow a large number of pages to be stored. The SRAM is connected to the address bus and data bus. The Serial Peripheral Interface (SPI) is used for communication with the Data Flash.

The Ethernet controller was originally a 16-bit ISA device, but can also be controlled in 8-bit mode. The Ethernet controller is configured as an 8-bit device. The Ethernet controller features 4K bytes of internal memory which is accessed through the I/O registers or directly through memory mapping of the entire memory. Default operation on the Ethernet controller is I/O mode and address 0300h. Since only address lines A0 - A12 are connected (need only 4 K bytes of address space), the I/O registers are mapped to address 8300h - 830Fh (I/O mode when the address lines into the PLD have the following configuration: bit 15 is high and bit 14 is low). By configuring the Ethernet controller through the I/O registers, the address can be changed and memory mode can be enabled. Memory mode operations can be mapped into address locations C000h - D000h.

III. SOFTWARE

The software running on the embedded web server follows the same layered structure as used in the TCP/IP protocol suite. The TCP/IP protocol suite

allows computers of all sizes, running different operating systems, to communicate with each other. It forms the basis for what is called the worldwide Internet; a Wide Area Network (WAN) of several million computers. The TCP/IP protocol suite is a combination of different protocols at various layers. Every layer acts independently from each other. An Ethernet controller driver controls the Ethernet interface. The Address Resolution Protocol (ARP) translates IP addresses to Ethernet MAC addresses (and vice versa) The Internet Protocol (IP) delivers packets to Transmission Control Protocol (TCP), UDP, and Internet Control Message Protocol (ICMP), the ICMP answers to PING requests and TCP/UDP delivers data to the applications. The applications can communicate with the transport layer through buffers with data and variables with control information. This section explains how the TCP/IP protocol suite is built up in our approach.

A. Link Layer

Data-link or Network Interface Layer is another common name of this layer. The Link Layer normally includes the device driver in the operating system and the corresponding Network interface (card) in the computer. Together they handle all the hardware details of physically interfacing with the cable. The Ethernet controller is configured to generate an interrupt every time a packet addressed directly to the Ethernet address arrives or when a broadcast arrives. When an interrupt occurs, the microcontroller reads the whole Ethernet frame into memory. A buffer of 1514 bytes, which is the maximum frame size on Ethernet, is reserved for this frame. Once the frame is transferred to the microcontroller, the Ethernet header is checked in order to ensure not receiving a misplaced frame. If the Ethernet address seen by the receiver is either a broadcast (all binary 1's) or addressed directly to specific Ethernet device, the frame is sent to the next layer or protocol according to the field, protocol type, in the Ethernet header.

B. Network Layer

This layer is sometimes called the Internet Layer. It handles the movements of packets around the network. Routing of packets, for example, takes place here. IP (Internet Protocol) and ICMP (Internet Control Message Protocol) provides the Network Layer in the TCP/IP Protocol Suite. The network layer controls the communication between hosts on the Ethernet. There is no form of transmission control to ensure that IP datagrams arrive to the host or that all IP datagrams from another host is received. This makes the layer rather easy to make. The ICMP sends messages between hosts and is only used to answer PING requests from a host. The IP handles communication for the overlaying Transport Layer.

C. Transport Layer

On the transport layer there are two major protocols which offer two different kinds of service; TCP which is a reliable delivery service and UDP which offers an unreliable service. TCP also offers flow control for retransmission of segments and acknowledgement of received segments.

D. Application Layer

The Application layer handles the details of a particular application. Several applications may be implemented in the embedded web server. The main limitation is memory usage and performance. Running several applications at once means lower performance.

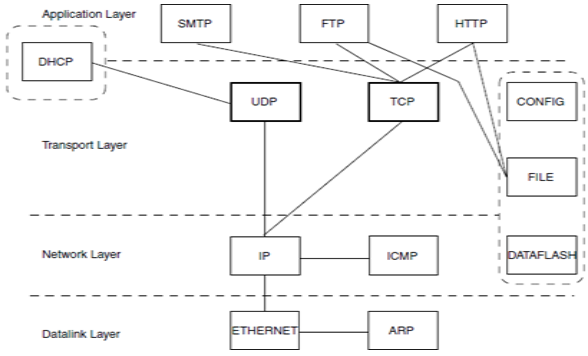
IV. PROTOCOL DEPENDENCIES

Figure shows the protocol dependencies for the modules in the Embedded Web server. The protocols are described below:

- HTTPD needs TCP and a file-system/Data Flash to operate.

- FTPD needs TCP and a file-system/Data Flash to operate.
- SMTP needs a running TCP implementation to operate.
- DHCP needs IP and UDP to operate. In the initialization phase DHCP requires that the IP and UDP protocol forward any IP packets delivered before the IP address is configured.
- TCP needs IP to operate.
- UDP needs IP to operate.
- ICMP needs IP to operate.
- IP needs ETHERNET to operate.
- ARP needs ETHERNET to operate
- ETHERNET depends only on the hardware Ethernet controller.
- CONFIG needs file-system/Data Flash to read the configuration file. If the configuration file is unavailable, it can return standard values.
- FILE needs DATAFLASH to work.
- DATAFLASH is only dependent on the hardware Data Flash.
- MAIN must initialize Ethernet, Data Flash, and file-system, TCP, DHCP and HTTPD if these protocols are to be used.
- DHCP, FTPD and HTTPD require repeatedly polling to operate.

User Developing Program



When an application sends data using TCP, the data is sent down the protocol stack, through each layer, until it is sent as a stream of bits across the network. Each layer adds information to the data by prepending headers and adding trailers to the data it receives.

Some abbreviations:

- TCP segment: The unit of data that TCP sends to IP.
- IP datagram: The unit of data that IP sends to the network interface.
- Frame: The stream of bits that flows across the Ethernet.

IP (Internet Protocol) adds an identifier to the IP header it generates to indicate which layer the data belongs to. IP handles this by storing an 8-bit value in its header called the protocol field. Similarly, many different applications can be using TCP or UDP at any time. The Transport Layer protocol stores an identifier in the header they generate to identify the application.

Both TCP and UDP use 16-bit port numbers to identify applications. The TCP and UDP store the source port number and the destination port number in their respective headers. The network interface sends and receives frames on behalf of IP, ARP, and RARP. There must be some form of identification in the Ethernet header indicating which network layer protocol generates the data. To handle this, there is a 16-bit frame type field in the Ethernet header.

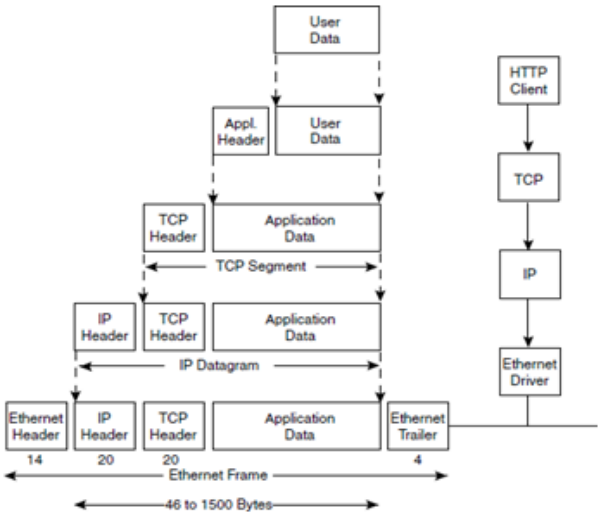


Fig.4 Encapsulation of Data as It Goes Down the Protocol Stack

Case Study-The web server apply to an Air Condition

The intention of this embedded internet server is hyperlink to tool and home equipment, the small size of permit that it is able to embedded into any little home equipment. we've got constructed several packages the use of this web server here we gift an example of utility that permit users to govern and display the kingdom of air-circumstance.

Embedded internet server connects with air-situation with I2C bus. On some other give up embedded net server connect Ethernet with RJ45 interface. The homepage of this application is set 28K byte in size; it consists of HTML files, thirteen images at jpg layout and Java Applet documents. pics are used to display the nation of air- situation vividly, Java Applet used to conversation among person and embedded web server. When consumer press the button on homepage of this air-circumstance corresponding command are send to embedded web server, among embedded internet server and browser we use UDP protocol. because UDP is an unreliable protocol so we reliability individual is guaranteed at Java Applet. while embedded net server obtained the command that person ship out from far flung browser it copes with this command consistent with the protocol with air-situation. I2C bus used between them. So, air-circumstance were given the command that far flung person ship. whilst air-situation complete the action then send lower back a reaction to embedded internet server shipping this response to remote person.

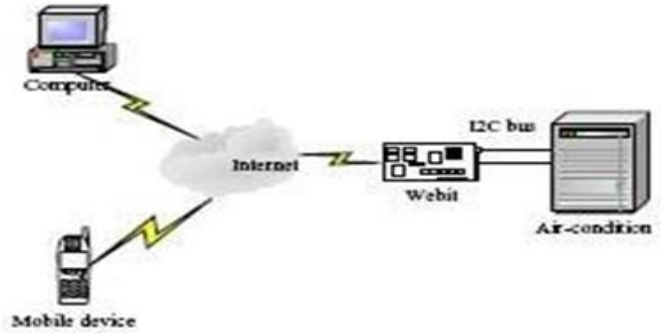


Fig.5 Controlling an air condition through Web server

V. CONCLUSION

we have delivered the overall design concept of the embedded net server and the coverage of TCP/IP discount, especial the reduction of TCP, whose aim is to permit clean get right of entry to and exploitation of far-flung gadget. This internet server gives the common devices a web interface and profits a terrific performance. it is able to be used widely in enterprise, clinical, and other fields, extra essential it may carry us a brand-new domestic existence. With the web server embedded, we are able to start to see the utility of computing technology in settings wherein they may be unusual these days device and appliance networking in the home; faithful capture of scientific experiments inside the laboratory; and automatic full-time monitoring of affected person.

VI.REFERENCES

[1]. Lexia Liu, Research on Technology of Embedded Web Server Application, IEEE, 2005.

[2]. Masato Shimano, Okazaki, Small Embedded web server for home appliances with embedded MPU & RTOS, IEEE, 2004.

[3]. Tao Lin, Hai Zhao, An Embedded Web Server for Equipment, IEEE, 2004.

[4]. [EME99] Gary Desrosiers, 99. Embedded Ethernet. http://www.embeddedethernet.com/ [Accessed Feb 2000].

[5]. [TCP86] Geoffrey H. Cooper. IMAGEN Corporation. http://www.csonline.net/bpaddock/tinytcp/default.htm [Accessed Feb 2000]

[6]. Behrouz A. Furazan (2006), Data Communication & networking, 4th Edition.

[7]. Raj Kamal (2004): Embedded System architecture, programming and design, 1st Edition.

[8]. Andrew S. Tanenbaum (2008): Computer Networks, 4th Edition.
[9]. Cirrus Logic Product Data Sheet, Crystal LANTM ISA Ethernet Controller.
[10]. Philips Semiconductors product data sheet: LPC2124.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Generalizing Machine Emotions

Aditya Purswani, Anirudh Pandita, Anasuya Dutta, Anusha Salimath
Student, Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

The goal of this research is to see if there is a way to create a universal machine that can represent emotion in any object. If effective, it might be utilized for a variety of purposes, including the development of EI software and the creation of situations involving EI robots by humans. It's not about being a soft when it comes to emotional intelligence! It's a new kind of intelligence: the ability to use one's emotions to aid decision-making in the time and gain more effective control over oneself and one's impact on others. Emotional Intelligence permits us to think more creatively and solve difficulties by utilizing our emotions. Emotional Intelligence and broad intelligence most likely coincide to some extent. A person who is emotionally intelligent is Identifying emotions, using emotions, comprehending emotions, and regulating emotions are all skills that you should have. The purpose of this paper is to review the literature on emotional intelligence (E.I.).

Keywords: Expression recognition, Convolutional Neural Network, Natural Language Processing, Multi-Level Perceptron,

I. INTRODUCTION

There are numerous human emotions that machines are capable of feeling, such as happiness, grief, and disgust. They differ in how they manifest themselves and what triggers them. The three most studied emotions in regards to the development of EI are happiness, anger, and sadness because they were found to be the ones that most prominently displayed empathy-related behaviours. For example, while a robot might be able to recognize another robot is sad by the way it is moving or looking down on the ground, it would not know what made it sad. A machine would only be able to detect the emotion, but not know what caused it. The authors were also interested in creating a machine that could sense other emotions in addition to happiness, anger, and sadness; they set out to explore what emotions would be most beneficial to creating such a generalized machine. They found that it would be beneficial for a machine to have the ability to recognize and express feelings of contempt, disgust, embarrassment, interest, joy, pride, relief, surprise and worry. They also thought that it would be beneficial for the machine to recognize and express all human facial expressions. They also believed that inorder for the generalized machine to be successful it would need a wide variety of triggers so that different situations could create different emotions. Due to the success of their test run, the authors decided to expand their study with a robot using automatic facial expression recognition.

II. WORKING

How exactly do computers recognize human emotions? Many of our modern computer systems use machine learning to identify a range of human emotions based on facial expressions, the tone and volume of voice, and even body language. Machine learning is a branch of artificial intelligence that allows computers to make inferences from data without being explicitly programmed. It's also used in various applications such as cybersecurity, internet bots, medical diagnosis, advertising fraud detection and more.

Machines can understand human emotions through the figure of speech of humans, their facial expressions and predict how it should react in order to generate positive feedback from the humans. This can be achieved by reinforcement learning

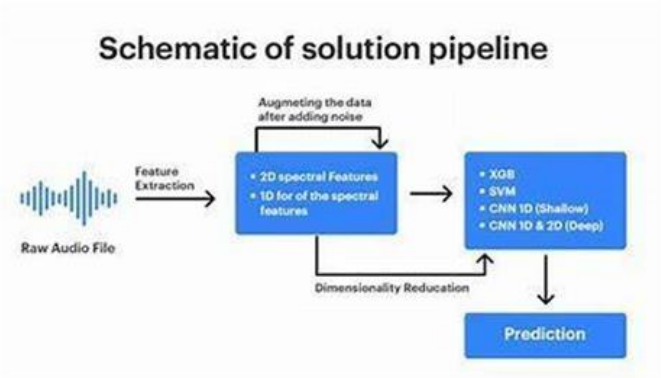


Figure 1: Schematic of solution pipeline

For example, if you've ever had a frustrating email exchange with an automated customer service chatbot that keeps repeating the same question over and over again despite not giving it any input whatsoever- well now you know why! The chatbot is using its data from previous conversations with humans to learn how to respond. One of the most common uses for machine learning is facial recognition. The industry uses a variety of algorithms to identify emotions using a sample of 100,000 images of people's faces. The accuracy depends on the specific algorithm utilized and the data set used (for example, if only pictures of women or only pictures with sunglasses are used then accuracy increases). Some companies have built entire facial-recognition applications that can identify what you're thinking or what you're feeling or even recognize your whole face with just one picture.

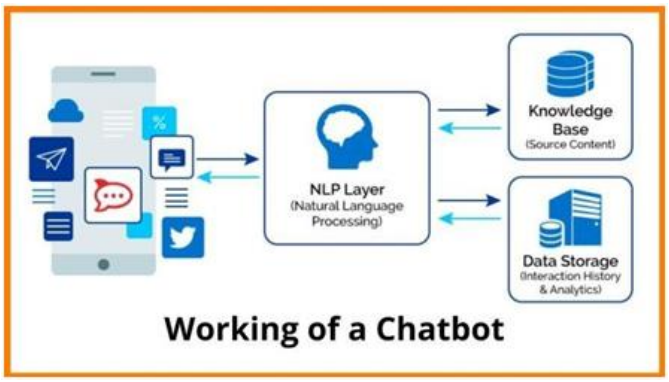


Figure 2: Working of a Chatbot

III. TECHNIQUES

Convolution Neural Networks

A CNN, or convolutional neural network, is a deep learning neural network designed to interpret structured arrays of data like photographs. Convolutional neural networks are widely utilised in computer vision and have become the state-of-the-art for many visual applications such as image classification, as well as natural language processing for text categorization.

Natural Language Processing

Natural language processing (NLP) is an artificial intelligence area that aids computers in comprehending, interpreting, and manipulating human language. In order to bridge the gap between human communication and machine understanding, NLP draws on a variety of fields, including computer science and computational linguistics.

Multi-Layer Perceptrons

MLPs are a type of neural network that is a subset of neural networks. One or more layers of neurons make

up these structures. The data is sent into the input layers, which may be followed by one or more hidden layers. The output layer is where the forecasts come from.

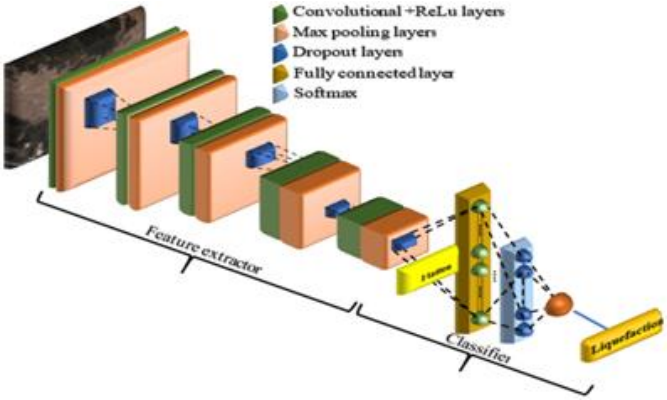


Figure 3:A Typical Convolution Neural Network

IV. APPLICATIONS

Sophia – The Humanoid

A humanoid robot named Sophia was activated for the first time on the 14th of February 2016 in Hong Kong. The robot is developed by Hanson Robotics and it has made its first appearance at SXSW, an American festival where people from different fields come together to talk about what they do. Sophia is a social humanoid robot that can exhibit natural facial expressions and can carry out simple conversations, both verbal and written. It is programmed to process speech and respond accordingly to its interlocutor.

In a video released by Hanson Robotics, Sophia's creator David Hanson introduces it as "a new form of technology that combines the traditional silicon-based computer with unique artificial intelligence software that allows it to build its knowledge and experience, like a true human". Currently, there have been many attempts to create humanoid robots, but currently, no technology has allowed them to become autonomous. However, Hanson Robotics does not consider Sophia as a mere robot but rather as artificial intelligence with consciousness and personality. Hanson Robotics has also introduced Jules, the first robot citizen of Saudi Arabia.

Figure 4: Sophia Internal Connections

Understanding Human Emotions in Hospitals

The term "motion intelligence" is a facet of emotional intelligence that has been used to describe one's ability to think about, perceive and sense their own body and its abilities. Here, we will discuss the application of motion intelligence for those living with disabilities.

In this post, we'll explore the idea that being able to move well is an essential part of quality patient care. We'll also explain how individuals in hospital settings can benefit from educating themselves on motion intelligence and discovering new ways to manage their physical environments.

Understanding Human Emotions in Traffic When a car crashes into another vehicle during an automobile accident, the person who caused the accident is typically in their own vehicle. They then have to be extricated from that car by emergency services. This is a process that can take hours or even days depending on the severity of the collision.

The traffic scene also has to be mitigated before traffic can return to normal, which may take hours. In this time, questions about what happened and what led up to it will inevitably arise for all parties involved.

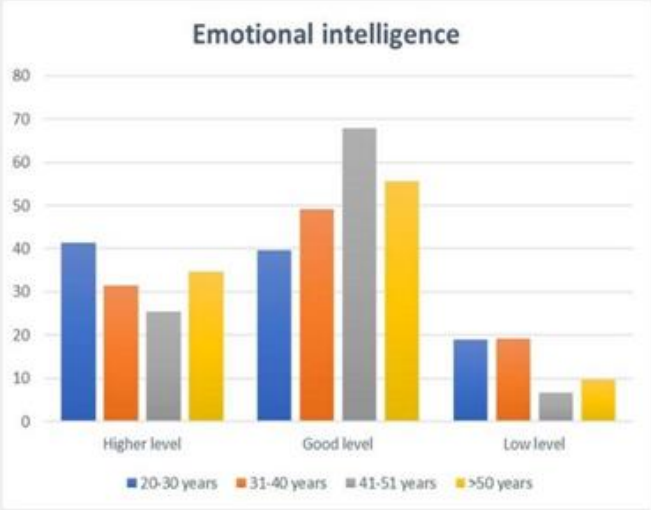


Figure 5: Levels of Emotional Intelligence in different age barriers.

Care for Lonely and Elderly People

We at times get to meet people that are elderly and lonely and have no one to look after them. Emotional Intelligence based robots can be used in order to help these people. There has been pioneering work in the field of computational psychiatry and neurophysiology that can enable us to understand how the brain works and what neural patterns are processed. These advances have allowed a substantial increase in the use of these techniques to help people.

One huge challenge for researchers hoping for a cure for mental problems is that there are so many things peoplecan do to harm themselves, like smoking, drinking or over-eating. However, using AI technologies can change this scenario by making it possible to create robots capable of understanding human emotion. However, AI has been used in many other fields in order to help people who are suffering from different mental problems. Over the past few years, we have seen that robots have been used as a form of treatment for people with chronic depression.

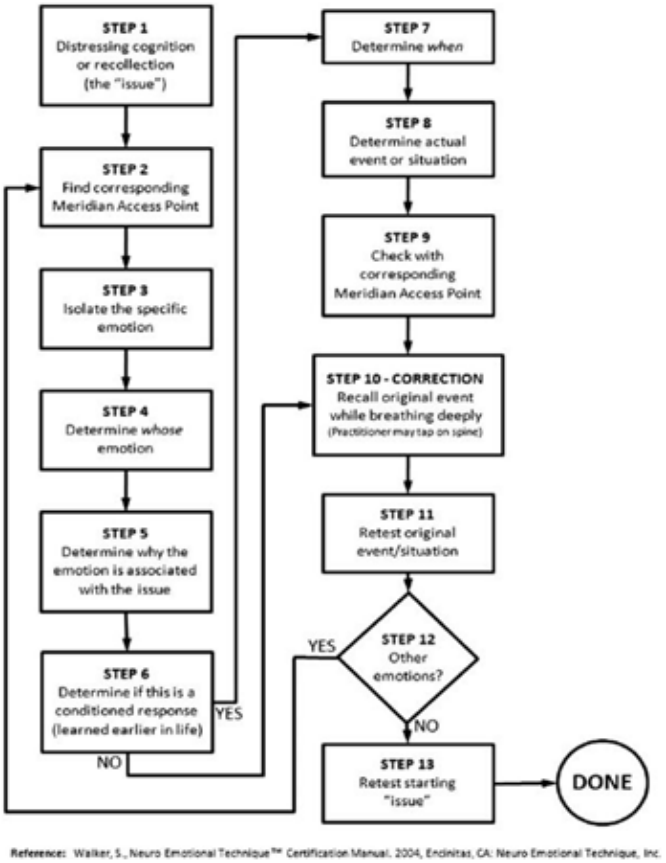


Figure 6: Working Flowchart of an EI Model achievement of life goals by mastering all the main concepts of emotional maturity.

How can someone develop Emotional Intelligence? It is essential to reflect and analyse oneself on a daily basis in order to achieve the maximum level of emotional intelligence. So, how can you assess yourself in order to improve your emotional intelligence? With the advancements in technology once can learn how to increase ones emotional intelligence and also apply the same techniques to machines so that they can learn to detect emotions and interact with humans more effectively and efficiently.

V. CONCLUSION

Emotional intelligence is a technique for people to think about their own feelings and how they want others to see them. Self-awareness, self-management, empathy, and social skills are the four basic components of emotional intelligence. Self-awareness allows us to reflect on our own actions. Having the time of day to manage ourselves and also the time to keep ourselves in check is self-management. Empathy is the ability of empathising with another' feelings. Building a stronger bond with others around you is a function of social skills. One can live a balanced life and assist in the

VI.REFERENCES

[1]. Alpaydin, E. (2014). Introduction to Machine Learning. Cambridge, Massachusetts: MIT Press.
[2]. https://www.researchgate.net/publication/285164623_An_Introduction_to_Convolutional_Neural_Networks
[3]. https://www.researchgate.net/publication/228340819_Multilayer_perceptron_and_neural_networks
[4]. https://en.wikipedia.org/wiki/Sophia_(robot)
[5]. Theory of Emotional Intelligence received fromhttps://positivepsychology.com/emotional-intelligence-theories/
[6]. The Measurement of Emotional Intelligence: A Critical Review of the Literature and Recommendations for Researchers and Practitioners https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6546921/
[7]. Natural Language Processing https://en.wikipedia.org/wiki/Natural_language_processing

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

User Grinding Using Machine Learning Algorithms

Bada Rajasekhar Reddy, Lakkisetty Naga Sai Neha, Boya Harshitha, V Bhanu Prakash
Department Computer Science and Engineering, New Horizon college of Engineering, Bangalore, Karnataka, India

ABSTRACT

In modern scenario, the clients of the corporation are getting extra attracted in the direction of predominant factors which can be provided by different way by the organization. The cutting-edge organizational technology is evidencing higher opposition in presenting present day technologically advanced services to the customers to hold them. Even though with efficient customer relationship management systems companies suffer from customer churn.In this paper the customer churn prediction using various machine learning algorithm is discussed. It can be more beneficial for the organization with customer attrition or churn model for maintaining better client relationships, increasing more clients and in improvement of the client retention by adding more profitable methods to the business. Machine learning gives different approaches such as logistic regression, decision trees, random forest, SVM algorithms used to predict the customer attrition can add more importance to the customer retention strategies. In this research paper the customer churn prediction models are analyzed with customer retention strategies.

Keywords—Machine learning, logistic regression, decision tree, random forest, KNN, SVM

I. INTRODUCTION

Customers are a constantly required and an important in increasing earnings and the revenue of the every organization subsequently to gain profit and increase the revenue. The customer's feedback about the organization is also very important. So the retention of customers is always plays an important position in lots of factors like growing income of the product and revenue of every corporation therefore, to advantage consumer pleasure it's far vital for the organizational managers and higher government to keep one rather effective and dynamic client relationship by using thinking about the relative features, goal clients and preserving effective relationship with the clients. In general higher the purchaser loyalty in the direction of the organization, lesser is the purchaser churn rate subsequently the usage of machine learning set of rules such as logistic regression, decision tree, random forest, k- nearest neighbor, support vector algorithm can predict and alert the organization about the condition of the churn rate. This paper will be focusing on the client churn prediction by the usage of various machine learning algorithms available and Client retention by the pattern in customer attrition and growing the retention rate.

The data is collected from the databases of the organization where it has all the details of the

customer. The data collected will be pre processed where data cleaning, missing data filling and other pre processing is done. The processed data will be given to the predictive model for the prediction. The model is trained with the train data, during the training the model learns the pattern and learns the customer behavior. The model predicts the outcome based on the learning from the training data.

II. BACKGROUND

Customer Churning is of two types either voluntarily or involuntarily. Voluntarily means the client is willingly makes a decision to terminate or discontinue the relations with a given supplier or the company. Involuntarily approach wherein the client's action or behavior renders them to their evacuation from assignment any form of commercial enterprise members of the organization with a given entity or business enterprise. Suddenly churning arises in instances where the purchaser does not obey to the stipulated set of rules and regulations that conducts commerce inside a given premise of thecompany,and for this condition it may be due to unlawful practices like refusing to pay timely or stealing.

Customer Attrition prediction is described as a supervised classification issue within the organizational area. The major or main focus of the model is to construct a predictive model to find the churn rate that learns the purchaser behavior and churn sample from features extracted from information exclusively to make an correct prediction forecast for the opportunity of no activity of the consumer associated with the employer in the future within predetermined churn duration. A churn prediction is based totally on a predefined time boundary set by using the model developer at which the duration starts and the given records are the functions which can be extracted from the statistics or database for every customer primarily based on the pattern previous to already described time boundary. The information of the clients after the time boundary is used to infer the prediction of the consumer as churn/non-churn labels for the classification purpose.

III. EXISTING WORK

Models primarily based classifiers like random forest, boosting and bagging can be compared with Support Vector machine, Decision tree and Naive Bayes Classifier the outcomes of the above models yielded that the ensemble- primarily based classifier had capabilities like low specificity, excessive sensitivity, extra accuracy much less error fee as compared to the other models. The models of SVM that are parametric and the parameters preliminary values could have a extremely good affect on their accuracy and performance of the model even while predicting the output. Deep neural networks (DNN) are also used in some cases to build a model of artificial neural network model that fits the information using various features and concepts to increase the overall performance of model constructed. This version has been trained with technique of k-fold move or cross validation approach for the training dataset. Many research specifically focuses on prediction model of the client churn by way of identifying the important thing elements which are essential, impacting the churn and which reason the churn. There has also been a hybrid or an amalgamation of different models with the usage of neural networks that has produced effects with a probability of 94%. Feature selection or extraction is done by the means of use of data benefit or statistics advantage and correlation characteristic by using rating filter out.

A churn score is mapped with every purchaser to locate any signs and symptoms of churn pattern of

their conduct. A predictive system getting to check the model suite of supervised algorithms are fed the very last feature dataset that's decided by way of making use of a brute force approach to extract the functions to do characteristic engineering, the usage of function choice strategies to become aware of the set of relevant attributes which contributes to the prediction of the churn score. There is a enormous use of churn prediction to find the retention of diverse social media customers by means of combining the social network engineering and contact log facts by way of generating a set of influential users who may have an impact on each other over the social media. A clear study has been conducted on the existing structures and methods, and flaws or disadvantages of every model had been recognized and after the analysis of the model, the high-quality or maximum suitable in all the aspects for the dataset has been chosen.

IV. DATASET

The dataset that associated with the paper work with the churn prediction of telecom customers, centered on a client or customer retention software in near future. The data in the each row inside the information set represents a client with capabilities or features as the attributes. The attributes of the customers are contained in each column. The feature set of churn column that depicts the records approximately about the clients who terminated their corresponding enterprise's offerings one month again. It also contains the information like offerings a given client opted for like net connectivity, coverage from damages, purchaser care services, number of cellular gadgets, security, variety of lines, streaming TV and movie services and backup plans. Add on to this it additionally consists of the information approximately a client's account details like the duration of the time for which they've accessed the corresponding services, their utilization plan, payment method that has been chosen for the services, monthly prices, billing service (papered or paperless) and general charges and ultimately the columns of the dataset of the features or informational data related to the demographic of the customers like gender, their marital fame, wide variety of kids and age. Feature choice could be carried out to put off some attributes which aren't wanted for the prediction and a few columns are dropped.

Data Preprocessing

Preprocessing of information is the method in machine learning in which transformation of raw data into an understandable format or required format for the model to process. Preprocessing of data is the process in machine learning of transforming raw data into an understandable format or required format for the model to process. It is also an important step in data mining as models cannot work with raw data. The quality of the data set must be checked and verified before applying machine learning or data mining algorithms. Data Preprocessing is mainly applied to check the data quality and their format. The quality can be checked by the following

- **Accuracy**: To verify whether the data given or entered is correct or not.
- **Completeness**: To verify whether the data is available in the data set or not recorded.
- **Consistency:** To verify whether the same data is kept in all places where ever it used, that do or do not match.
- **Timeliness**: The data must be updated correctly.
- **Believability**: The data should be trustable and valid.
- **Interpretability**: The understand ability and the ease of understanding the data.

Major Tasks in Data Pre Processing

- Cleaning of Data
- Integration of Data
- Data Reduction
- Transformation of Data

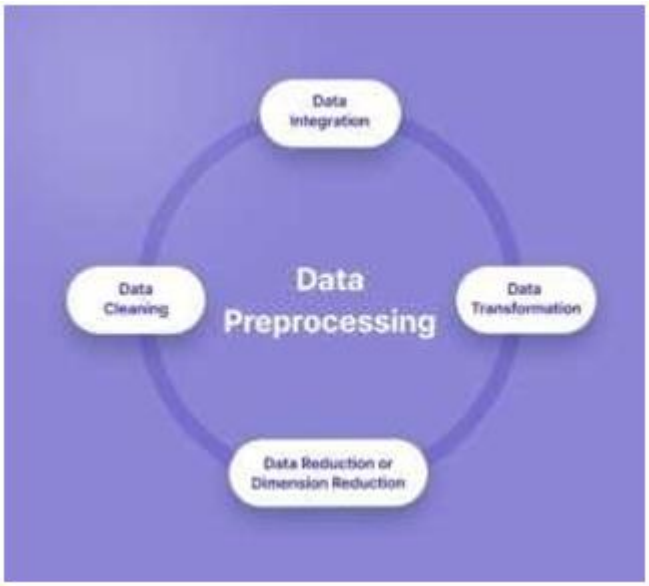


Fig 2. Data Preprocessing

V. MACHINE LEARNING ALGORITHMS

Gradient Descent

Gradient Descent set of rules is an iterative method wherein the main goal of the model is to limit a cost characteristic that is carried out in the model. The model construct must to be able to compute the partial by-product of the feature which is also referred to as slope or gradient. The coefficients concerned inside the problem are calculated at each of the iteration by means of considering the bad cost of the derivative and by decreasing the coefficients at each step of behavior learning with the consideration by a learning rate (step size) extended via spinoff in order that the local minima is accomplished after a few iterations inside the model. So the iterations are stopped whilst gradient converges to minimal cost of the value function after which there's no further minimization in value function. There are specifically 3 special forms of gradient descent approach: "Mini Batch Gradient Descent" (MBGD), "Batch Gradient Descent"(BGD) and "Stochastic Gradient Descent" (SGD).

Linear Regression

Regression method comes under the supervised learning approach. It is specifically used to train a model with variables which are not continuous and do the predictions. The following are the real time application example of linear regression algorithm such as prediction of the exam scores of school students, forecasting of moves, in the price of inventory in inventory change and prediction of charge of actual- property, forecasting of sales. In Regression models the problems are solved with the usage of regression could have the categorized datasets and the output variable cost or anticipated value can be decided by means of input variable values or given variable values so it is the supervised learning method in which the version known's the output variables. The most easy type of regression methods is linear regression in which the strive is made to healthy a instantly line such as instantly hyper plane over the dataset and it is viable best if while the relationship between the target and entered variables of dataset is linear in nature. Linear regression has benefit that it is straightforward to recognize the model and it's also clean to keep away from the over fitting via regularization. Linear Regression is a good approach if it is known form dataset that the relationship between response variable and covariates is linear. The model shifts the focal point from statistical modeling to records analysis and preprocessing. Linear Regression is good and simple approach for learning about the statistics evaluation procedure. However, it is not a advocated technique or technique for most practical programs as it oversimplifies the real global problems in standard.

A linear regression simple version of model has variables namely the target variable guided with the aid of an impartial or input variable. But in actual existence it is able to not be the case as they are more complicated and depending on multiple impartial variable or characteristic. In general one target variable depends on more than one elements. For instance, the residence fee of a residence relies upon on many factors and features just like the community it is situated in and round, location of it, wide variety of rooms in the residence, location of nearest airport / station from it, attached centers with services, location of nearest purchasing from it, etc. In simple linear regression the variables will be having a 1-1 relationship between them. Whereas in multiple linear regression model, there may be a M-1 relationship, between some of independent predictor variables and one dependent response variable. Addition of extra or more than one predictor variables to the model will now not constantly mean that the regression can be higher in the performance and accuracy, or there is a better improvement in the predictions. Simple and multiple linear regression have their own exclusive use instances in real time issues and one is not superior than the alternative. In a few use case problems including extra variables can make the prediction worse than earlier than because itoutcomes in over- fitting the dataset over the model. Again as greater input variables are brought in the model it creates extra relationships among them which leads to more complexity to the model to expect the output fee. So now not only the enter or predictor variables doubtlessly associated with the output or response variable, they're additionally potentially related and have an effect on to each different within the prediction, this scenario is also called multicollinearity in ML domain. The optimum solution is in which all of the enter variables to be correlated with the response variable, but now not with every other.

Logistic Regression

Logistic regression is used in the situation where the statement deals with a classification of the enter information into some label outputs. The logistic regression models offers the probability of the outcome as if an event will occur or not which can be also in terms of 0 or 1 based on data provided as variables. For example, an electronic mail from a supply is assessed as junk mail or not, predicting whether a tumor present in human is malignant or benign based on the use of features are the times which can be taken for the binomial prediction or classification of Logistic Regression. There also can be multinomial final results of Logistic Regression model as proper prediction of sort of cuisine desired like Mexican, Italian, Chinese, and lots of others. There may be ordinal outcome of the problem like product score 1 to 5 and plenty of others. So Logistic Regression deals with prediction of goal variable that's specific and classified in nature. Where in simple regression the model does the prediction of values which can be continuous e.g. Residence charge prediction of real estate houses over a span of three years with the usage of many corresponding capabilities and information.

Decision Tree

Decision Tree method is the Supervised Machine Learning method used in solving a type and regression problems by means of splitting continuously the input records based totally on a particular parameter with the intention to have powerful influence within the model prediction. The structure of the decision tree are in such a way that all the decision are in the leaves node of the decision tree and the records are present in divides in the nodes. The decision tree

Classification variable is categorical (final results which is in the form of No/Yes) and the decision variable in the regression is continuous. Decision Tree has the following benefits as it's far suitable for classification as well as regression problem which can be modelled by this type of algorithm, it makes easy in understanding the manipulating the quantitative and explicit values ,data, helps to fill the missed attribute values with most appropriate value ,the efficiency set of rules of tree traversal helps to increase overall performance The Decision Tree method may come across the over-fitting problem therefore, ensemble modeling method in Random Forest is the best solution.

SVM

Support Vector Machines (SVM) is one of the ancient and also a model which could handle one is classification and other is regression problems very efficiently. In this method hyperplane as it's parameter requires to be describe which is the decision boundary of the model. When there are a set of objects be owned to various classes then already decision plane is used to divide them into distinct classes of objects. The objects shall or shall not be linearly separatable in this model however if the use case complexity is more than the a mathematical functions then called as kernels are implemented in order to divide the objects from the data which are members of different classes. SVM focuses at correctly differentiating the objects based on the given data examples like in the training data set. Following are the few advantages of SVM, it can manage both structured data and semi structured , it can even manage elaborate function problems if the model is set with proper and appropriate kernel function. The SVM models there is less probability or less chances of over fitting the data over the model is done using model generalization. It could also be scaled up with proper functions with high dimensional data. It will not have difficulty in local optimal.

Bayesian learning

Bayesian Learning is mainly comes under a prior probability distribution which is first chosen and then changed or updated to get a posterior distribution. Based upon the obtainability of new consideration of the earlier posterior distribution can be used as a prior in the model. Incomplete datasets can also be handled by Bayesian network model which can also handle the incomplete datasets. This method could be prevented over- fitting of data. It is not compulsory need to remove Inconsistence from the data which would be handled by the model on itself. Bayesian Learning has the following some of the disadvantages such as very difficult in selection of prior and has to be done correctly. Posterior distribution could be effected by prior to a great extent which might be negative as well as positive. If the prior chosen one is not correct which will lead the way to wrong predictions and the accuracy of model will drop in or weaken. It can be expected expensive and intensive. Bayesian Learning is used for the applications like victim recognition and medical diagnosis of the disaster and disease etc.

To overcome the issues in Bayesian studying Naïve Bayes method is followed. Conditional probability used in this algorithm to solve easily. The probable values in the table are used in this model through training of the data which will be updated during the training of the model. The feature values that are present in the probability table are used during training of the data and new tables with new values has to be recorded from the tables that already existed. The conditional dependence is used in this model hence it is called as “naïve”. All the features of the input are independent to each other that rarely becomes true because it does not become true always

in practical world. This algorithm has so many advantages as it is simple, good performance, easy implementation, it works good with the data which trained less, it works well with wide variety of records points and predicators, handles the records that are discrete and continuous, can take care of multi- elegance classification troubles and binary, which makes predictions that are probabilistic. This algorithm handles discrete and continuous records. This algorithm is not good for features that are irrelevant.

K Nearest Neighbor

K Nearest Neighbor also known as KNN algorithm is a type set of rules that are used for categorization of records. It carries a database of the dataset that is having information factors grouped or clustered into several groups, this algorithm will categorizes the given sample information point to it as a classification. KNN does no longer assume or make assumptions as any underlying records distribution and for this feature the algorithm is also known as the non-parametric. Some of the advantages of KNN set of rules are the subsequent it is straightforward approach that may be effortlessly carried out. Building the model is reasonably- priced and cost powerful. It is very flexible in class scheme and well perfect for issues having Multi-modal classes. Data records that are having the multiple magnificence labels. Error from this model is at most two times the other models like Bayes mistakes rate. It can be the fine method or best suited for some problem. KNN overall performance is better compared to SVM in the protein feature prediction with the use of expression profiles.

K Means Clustering

K Means Algorithm or K means Cluster is most regularly used for fixing problems concerning grouping or cluster of the statistics data. It is approach under the unsupervised learning in which it does now not have any categorized statistics data. It contains the subsequent benefits, it's far computationally greater than the normal hierarchical clustering while features are massive. K Means with other clusters like globular and small k it produces the clusters that very close to each other than different clustering. Easy to implement and interpretation of the clustering effects are the major attraction of this set of rules. The complexity of the K means in order to train is O(K*n*d).

VI. RESULTS

The dataset is provided to the data preprocessing process where all the records in the dataset is made in model processable format. The preprocessed dataset then undergoes the feature extraction where all the irrelevant features are removed and only needed feature are given to the model. The model uses any of the above mentioned machine learning algorithms to predict the churn rate. The existing modelsuses one or hybrid of two or more machine learning models to prediction. The customer churn prediction on various sector like telecom, banking, Saas platform are found in the various models using machine learning algorithms with different methods of customer churn prediction. The results of various machine learning models are given below.

TABLE I Customer churn Telecom

S.N	*Models*	*Train Data*	*Test Data*
1	RIDGE	95.92	66.28
2	RF	89.69	51.26
3	KNN	67.04	91.80
4	Deep Neural Network	67.04	61.84
5	SVC	44.85	37.29

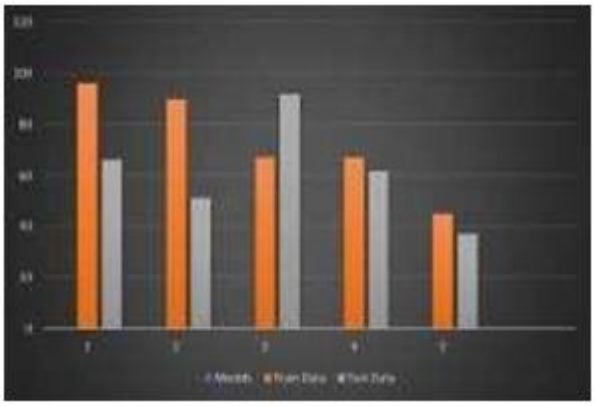


TABLE II Customer churn in banking

S. N	RESULTS BY CLASSIFICATION DIRECTLY		
	Models	*Accuracy(%)*	*Accuracy(%) Accuracy After oversampling(%)*
1	RF	86.19	94.75
2	DT	88.99	92.99
3	SVM	87.99	60.36
4	KNN	71.65	91.37

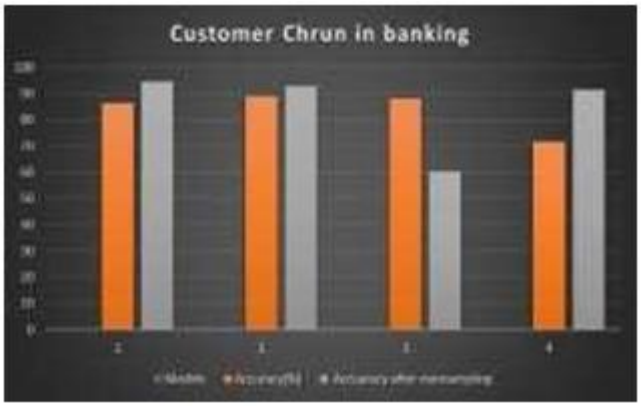


TABLE III Customer churn in banking

S. N	RESULTS BY FEATURE SELECTION		
	Models	*Accuracy(%)*	*Accuracy(%) Accuracy After oversampling(%)*
1	RF	88.55	43.18
2	DT	72.61	70.74
3	SVM	70.63	59.53
4	KNN	91.15	70.99

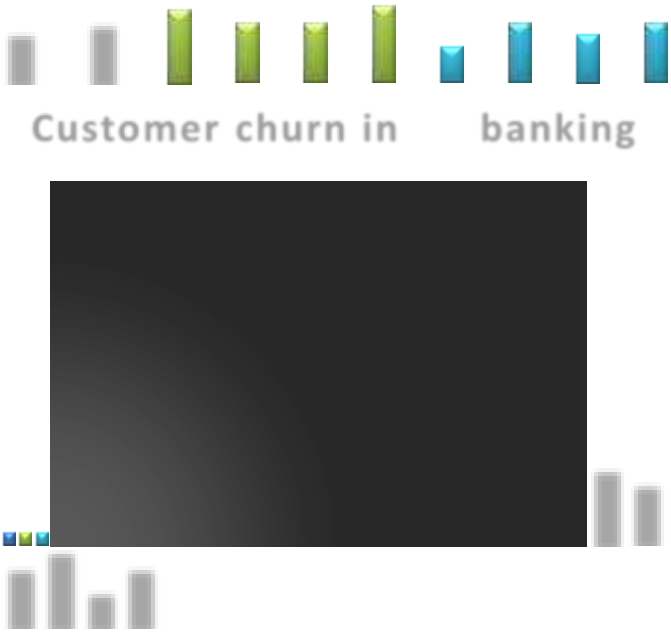


TABLE IV Customer churn in Saas Sector

VII. CONCLUSION

Organizational sector in today is packed with the digitalized statistics it performs a critical position in terms of the industries to have a quality survival within the international standards. In that situation the Client prediction of churn is very important and the good way in locating the client behavior. The churn prediction classifies clients into churn and non-churn clients based on their buy and the corresponding capabilities taken in attention. The datasets are mostly accrued from the churn data of the organization. The accumulated records is first normalized after which they are given to the model and looked after. The facts then given to a model getting to know prediction and the churn is expected. This paper offers the information about the models used for the customer churn prediction.

VIII. REFERENCES

[1]. S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on

Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[2]. P. Amornvetchayakul and N. Phumchusri, "Customer Churn Prediction for a Software-as-a-Service Inventory Management Software Company: A Case Study in Thailand," 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), 2020, pp. 514-518, doi: 10.1109/ICIEA49774.2020.9102099.

[3]. L. Butgereit, "Big Data and Machine Learning for Forestalling Customer Churn Using Hybrid Software," 2020 Conference on Information Communications Technology and Society (ICTAS), 2020, pp. 1-4, doi: 10.1109/ICTAS47918.2020.233972.

[4]. S. Suresh, D. Visvalingam, A. Lu and B. Wright, "Evaluating and Improving Attrition Models for the Retail Banking Industry," 2020 Systems and Information Engineering Design Symposium (SIEDS), 2020, pp. 1-6, doi: 10.1109/SIEDS49339.2020.9106629.

[5]. P. Sadana and D. Munnuru, "Machine Learning Model to Predict Work Force Attrition," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-6, doi: 10.1109/I2CT51068.2021.9418140.

[6]. N. Kolli and N. Balakrishnan, "Hybrid Features for Churn Prediction in Mobile Telecom Networks with Data Constraints," 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2020, pp. 734-741, doi: 10.1109/ASONAM49781.2020.9381482.

[7]. R. K. VLN and P. Deeplakshmi, "Dynamic Churn Prediction using Machine Learning Algorithms - Predict your customer through customer behaviour," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1- 6, doi: 10.1109/ICCCI50826.2021.9402369.

[8]. M. Rahman and V. Kumar, "Machine Learning Based Customer Churn Prediction In Banking," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1196-1201, doi: 10.1109/ICECA49313.2020.9297529.

[9]. P. Bhuse, A. Gandhi, P. Meswani, R. Muni and N. Katre, "Machine Learning Based Telecom-Customer Churn Prediction," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 1297-1301, doi: 10.1109/ICISS49785.2020.9315951.

[10]. V. Geetha, A. Punitha, A. Nandhini, T. Nandhini, S. Shakila and R. Sushmitha, "Customer Churn Prediction In Telecommunication Industry Using Random Forest Classifier," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262288.

[11]. L. Butgereit, "Work Towards Using Micro-services to Build a Data Pipeline for Machine Learning Applications: A Case Study in Predicting Customer Churn," 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), 2020, pp. 87-91, doi: 10.1109/ITCE48509.2020.9047807.

[12]. L. Wu and M. Li, "Applying the CG-logistic Regression Method to Predict the Customer Churn Problem," 2018 5th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS), 2018, pp. 1-5, doi: 10.1109/IEIS.2018.8597855.

[13]. B. N. K. Sai and T. Sasikala, "Predictive

[14]. Analysis and Modeling of Customer Churn in Telecom using Machine Learning Technique," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 6- 11, doi: 10.1109/ICOEI.2019.8862625.

[15]. X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 129-132, doi: 10.1109/ICCCBDA49378.2020.9095611.

[16]. H. Zhang, Z. Li, H. Shahriar, L. Tao, P. Bhattacharya and Y. Qian, "Improving Prediction Accuracy for Logistic Regression on Imbalanced Datasets," 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC),2019, pp. 918-919,doi:1109/COMPSAC.2019.00140.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Artificial Intelligence and Neural Networks for Probabilistic Reasoning

Charitha P Reddy[1], Ananya B Pinglay[1], Bashrath Tanzim[1], Dr. M Nirmala[2]

[1]UG student, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Associate Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Artificial Intelligence is the branch of Computer Science that aims to create the intelligence of the machines. Artificial Intelligence has a greater impact on the world. Large amounts of data are being produced by varieties of applications and other sources. It is a mixture of computer science, pure mathematics and data analytics [1]. Machine learning is the main constituent of Artificial intelligence, and it deals with the learning process and input data. So Artificial Intelligence is the perception, manipulation, reasoning, communication, and learning [3]. Artificial Intelligence has been implemented in an extensive range of fields like medical, robotics, law, scientific discoveries, and toys.

Key words: Machine learning, perception, reasoning, communication and learning

I. INTRODUCTION

Artificial Intelligence which is also called as Machine Intelligence is the intelligence exhibited by machines in comparison to the natural languages of humans and other animals [2]. Speech (voice) recognition, learning ways, planning and problem solving are some of the activities that are designed to be done.

Fig No.1 Artificial Intelligence

Robotics is the field which is concerned with perceptron to action connection [1]. So, Artificial intelligence plays a major role in Robotics. Artificial Intelligence addresses some of the pivotal questions like how the knowledge should be represented, how the knowledge should be used, what knowledge is necessary in thinking aspects. Robotics demands Artificial Intelligence by making it deal with real objects of the world [1]. The definition of Artificial Intelligence is not dependent on the time.

The success of Artificial Intelligence was due the magnificent power of the computers today, a prominent emphasis on solving certain subproblems, the formation of new bond between Artificial Intelligence and the other fields which are working

on similar kind of problems, and above all of these the dedication of researchers [4].

II. TOOLS

Artificial Intelligence has led to the development of huge variety of tools which are used to solve most of the difficult problems in computer science. Some of these methods are:

2.1 Search and optimization

Most of the problems in Artificial intelligence can be solved by searching through the problems in an intelligent way. By performing a search, reasoning can be reduced. For example, logical proof can be looked upon as finding a path where inference rule is applied in each step [2]. Local searches in configuration space are used by Robotics algorithms for grasping of objects and for movement of limbs [2].
Most of the learning algorithms make use of search algorithms which are based on optimization. For most of the problems, we can possibly begin the search from a guess and then clarify the guess gradually until no more clarifications can be made [3]. Stimulated annealing, random optimization and beam search are some of the optimization algorithms.

2.2 Logic

For representation of knowledge and problem solving we use logic, which can also be applied to other problems. For example, the satplan algorithm make use of planning and also for inductive logic programming. Fuzzy logic is the first – order logic where the truth of a statement can be represented as a value in between 0 and 1 [4]. Default logics, circumscription and non-monotonic logics are different forms of logic which are designed to help with default reasoning [2].

III. DESIGN AND IMPLEMENTATION

Many algorithms like Linear Regression, Logistic Regression, Decision Tree, Naïve Bayes Algorithm, Support Vector Machine Algorithm are used in Artificial Intelligence.
Linear Regression is one of the most widely used algorithms in Machine Learning [7]. To predict the output variables (future values), a significant variable from the dataset is chosen. If the labels are continuous, like the number of flights daily from an airport, Linear Regression is used in such cases.
The Representation of Linear Regression is $y = b*x + c$. According to the above representation, 'y' is an independent variable, whereas 'x' is the dependent variable. When the Linear Regression is plotted, then the slope of the line is termed as 'b' which provides us the output variables and 'c' is its intercept. The linear regression algorithms assume that there exists a relationship between the input and the output. There will be a loss in output if the dependent variables and independent variables are not plotted in linear regression on the same line. The loss in output in linear regression can be calculated using loss function [8].
Loss Function: $(\text{Predicted output} - \text{Actual output})^2$

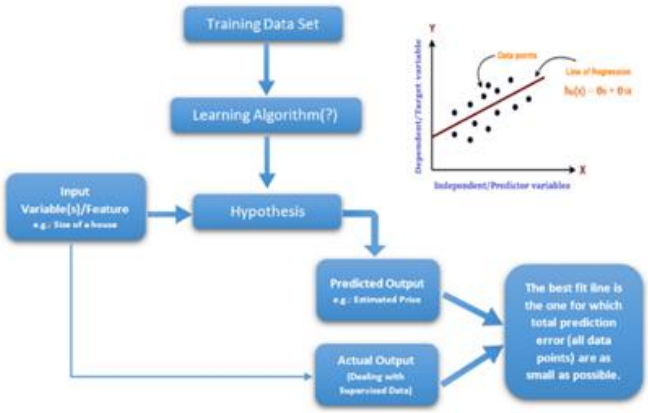


Fig No. 2 Linear Regression Algorithm detailed view

Linear regression can be implemented in several ways, some of them are: Sales prognostication, Risk analysis, Housing applications, Finance applications

IV. PROBABILISTIC METHODS FOR UNCERTAIN REASONING

Most of the problems in Artificial Intelligence need a medium to operate uncertain information or incomplete information. Artificial Intelligence researchers have come up with a number of potent tools for solving these problems by using probability theory methods and economics methods [6].

"Utility" is the key concept from the science of economics which is a measure of how precious a thing is to an intelligent agent.

Explicit mathematical tools have been advanced to analyse how an agent can use decision theory, decision analysis and information evaluation theory for making choices and planning [2].

V. CLASSIFIERS AND STATISTICAL LEARNING METHODS

Artificial intelligence applications can be divided into two types: classifiers and controllers. Controllers segregate the necessary conditions before deducing actions, and hence classification forms a core part of many Artificial Intelligence systems [8].

Classifiers are the functions that use pattern matching to ascertain the nearest match. They can be adjusted according to the examples, which makes them very appealing for use in Artificial Intelligence. These examples are called as observations or patterns [7]. A pattern belongs to particular predefined class in supervised learning. A class can be viewed as a decision to be made. Data set is a combination of all the observations along with their class labels.

An observation is analysed based on previous experience when a new observation is obtained. There are various ways to train a classifier such as by statistical and by machine learning approaches. Neural network, kernel methods like support vector machine, k-nearest neighbor algorithm, Gaussian mixture model, naïve Bayes classifier and decision tree are the most frequently used classifiers [8].

The performance of these classifiers have been contemplated over a vast range of tasks [2]. The performance of the classifier relies upon the characteristics of the data which is need to be classified. Determining the best classifier for a given problem is an art more than science.

VI. NEURAL NETWORKS

Neural network is based on the design and functioning of human brain and components. It is a series of algorithms that endeavours to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates [7]. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.

A neural network works similarly to the human brain's neural network. A "neuron" in a neural network is a mathematical function that collects and classifies information according to a specific architecture. The network bears a strong resemblance to statistical methods such as curve fitting and regression analysis [7].

Fig No.3 Artificial Neural Network

There are three main components in Neural Networks: an input later, a processing layer, and an output layer. The inputs may be weighted based on various criteria [8]. Within the processing layer, which is hidden from view, there are nodes and connections between these nodes, meant to be analogous to the neurons and synapses in an animal brain.

Types of neural networks

- Perceptron
- Feed Forward Neural Network
- Multilayer Perceptron
- Convolutional Neural Network
- Radial Basis Functional Neural Network
- Recurrent Neural Network
- LSTM – Long Short-Term Memory
- Sequence to Sequence Models
- Modular Neural Network

VII. FUZZY LOGIC

Fuzzy logic is a method of reasoning that is similar to human reasoning [4]. The approach of fuzzy logic resembles the way of decision making in humans which involves the intermediate possibilities between the digital values YES and NO.

The fuzzy logic works on various possibility levels to achieve a definite output. It can be implemented in systems in the range from small micro-controllers to large, networked, workstation- based control systems which can be of various sizes and capabilities [6].

7.1 Fuzzy logic in AI

In artificial intelligence (AI) systems, fuzzy logic is used to imitate human reasoning and cognition. Rather than using strict binary cases of truth, fuzzy logic involves 0 and 1 as extreme cases of truth but with several intermediate degrees of truth [5].

As a result of this, fuzzy logic is better suitable for the following: engineering for decisions without clear certainties and uncertainties, or with imprecise data such as with natural language processing technologies; and regulating and controlling machine outputs, according to multiple inputs/input variables such as with temperature control systems. IBM's Watson supercomputer is one of the best examples of how alterations of fuzzy logic and fuzzy semantics are being used [4].

VIII. APPLICATIONS OF ARTIFICIAL INTELLIGENCE

Artificial Intelligence has various applications in today's society. It is becoming essential for today's time because it can solve various complex problems with an efficient way in multiple industries, such as Healthcare, entertainment, finance, education, etc.

Fig No.4 AI Applications

8.1 Health care

In the last, five to ten years, AI is becoming more beneficial for the healthcare industry and going to have a notable impact on this industry.

Healthcare Industries are making use of AI to make a diagnosis better and faster than humans. AI can help doctors with diagnoses and can inform when patients are worsening so that medical help can reach out to the patient before he/she is hospitalized.

8.2 Finance

AI and finance industries are the best matches for each other. The finance industry implements automation, chatbot, adaptive intelligence, algorithm trading, and machine learning into financial processes [2].

8.3 Data Security

The security of data is important for every company and cyber-attacks are increasing in a rapid way in the digital world. AI can be used to make your data more protected and safe. Some examples such as AEG bot, AI2 Platform are used to ascertain software bugs and cyber-attacks in a better way.

8.4 Social Media

Social Media sites such as Facebook, Twitter, and Snapchat contain billions of user profiles, which are need to be stored and managed in a well-ordered way. AI can categorize and manage huge amounts of data [3]. AI can analyse lots of data in order to identify the latest trends, hashtag, and requirement of different users.

8.5 Robotics

Artificial Intelligence has a exceptional role in Robotics. Usually, general robots are programmed in such a way that they can perform repetitive tasks, but with the help of AI, we can produce intelligent robots which can carry out tasks with their own experiences without being pre-programmed [2].

Humanoid Robots are best-suited examples for AI in robotics, recently the intelligent Humanoid robot named as Erica and Sophia has been developed which can talk and behave like humans.

8.6 Agriculture

Agriculture is an area which needs various resources, labour, money, and time for its best output. Now a day's agriculture is becoming digital, and AI is making appearance in this field. Agriculture is applying AI as agriculture robotics, solid and crop monitoring, predictive analysis [3]. AI in agriculture can be very useful for farmers.

8.7 Education

AI can automate grading system so that teachers can have more time to teach. AI chatbot can communicate with students as a teaching assistant. In the future, AI can work as a personal virtual tutor for students, which can be accessed easily at any time and any place.

8.8 E – commerce

AI is providing a fierce competition to the e-commerce industry, and it is becoming more challenging in the e-commerce business. AI is helping customers to discover related products with recommended size, color, or even brand.

8.9 Travel and Transport

AI is becoming highly demanding for travel industries. AI is capable of doing various travel related tasks such as from making travel arrangement to suggesting the hotels, flights, and best routes to the customers [8]. Travel industries are using AI-powered chatbots which can make human-like interaction with customers for better and fast response.

8.10 Astronomy

Artificial Intelligence is very helpful to solve various complex universe problems. AI technology can be helpful for understanding the universe such as how it works, origin, etc.

IX. RESULTS AND DISCUSSION

The boundaries of machine-enabled functionalities are being pushed by Artificial Intelligence. It results in effective execution of iterative tasks. AI reduces the time taken to perform a task and also performs multi-tasking and eases the workload for the existing resources. AI has no downtime and operates 24/7 without any interruption. AI can be deployed across industries as it has mass market potential. By making processes faster and easier, AI facilitates decision-making. AI reduces human errors and increases accuracy and precision. AI has helped us come up with new inventions in almost every field to solve complex problems.

X. CONCLUSION

Artificial Intelligence is the kind of change which we certainly should not take it for granted. It is the change which not only starts to exhibit its positive impact soon on the society but severely negative impacts too. Although advances are likely to improve the functioning of AI, AI will remain a function of human activity. However, if AI can learn to self-replicate and thus become a life form, albeit a man-made one, outcomes become uncertain.

XI. REFERENCES

[1]. Prof. Allen, J. F. (1984). Towards a general theory of action and time. Artificial Intelligence,

[2]. George Luger, Artificial Intelligence: Structures and Strategies for Complex Problem Solving, Fourth Edition Addison-Wesley, 2002 -- a well-respected introduction to artificial intelligence, as witnessed by its being in its fourth edition.

[3]. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Second Edition, Prentice-Hall, 2003 -- the leading introductory textbook in the field.

[4]. Nils J. Nilsson, Artificial Intelligence: A New Synthesis, Morgan Kaufmann Publishers, 1998 -- Another fine introductory textbook on artificial intelligence.

[5]. Gary Riley, CLIPS: A Tool for Building Expert Systems, 2002

[6]. American Association for Artificial Intelligence (AAAI), Welcome to AI Topics, 2003

[7]. Artificial Intelligence: The Basics by Kevin Warwick, 2011

[8]. Intelligence Unbound: The Future of Uploaded and Machine Minds by Russell Blackford , 2014

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Detailed Survey on Internet Radio Multicasting over IP

Narhare Swati M D, Manikandakumar M, Salangai Nayagi, Devi Naveen
Department of Computer Science and Engineering, New Horizon College of Engineering, VTU, Bengaluru, Karnataka, India

ABSTRACT

Using mobiles, laptops, television and other devices we've come to be extra reliant at the vedio streaming services for entertainment. the internet's ever-increasing services and applications have dramatically increased the global inter-network. individuals all over the global for conversation(emails, text, messages, etc.) leisure(social networks, chat rooms, blogs, and many others.) And enjoyment(video games, literature, track, films, buying, and so on.)and exchanging statistics(schooling, geographic records, and many others.)this assignment has the capacity to be utilized for internet radio in addition to live video streaming. youtube live, facebook stay and different video conferencing and streaming offerings have grown in popularity to the factor that they now supply heaps of stay vedio streams to hundreds of thousands of visitors at any given time. For stay video distribution, content material vendors use unicast mode. Multicast inside the cell radio get admission to community(RAN)is potential option for correctly addressing he growing desires for vedio traffic.

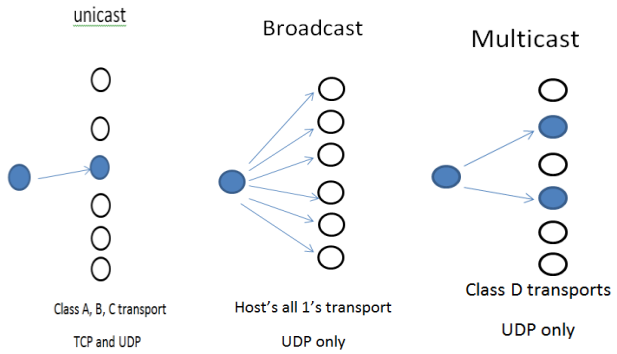
Keywords: Globalization of the media, media industries, media infrastructure, outsourcing, streaming tv.

I. INTRODUCTION

A one–to–many records distribution model is needed for radio and television broadcast programs, in which information travels. Unicast transmissions are one-to-one connections in which each user sends and receives data separately. Unicast broadcasting has a number of advantages. that each transmission may be controlled to a fine granularity separately. As a result, service providers will be able to They might provide customized unicast services, as well as on-demand services. For their customers, they provide demand and time-shifted watching options. which are not bound by service costs In unicast broadcasts, from a network standpoint, When a person is online, network resources are spent. from a one sender to multiple receivers at the same time, but not throughout the entire sub network. IP multicasting is used for this. IP multicast allows data to be sent from a sender to a group of receivers. It's an internet television or radio application in which the client opens a control channel with a TCP (transmission control protocol) connection and receives multimedia streams using a UDP (user datagram protocol) socket. To execute these actions, the user can pause, resume, adjust, and even terminate the channel, and a user-interactive GUI (graphical user interface) based program is created. Unicast transmissions are one-to-one connections in which each user sends and receives data separately. Unicast broadcasting has a number of advantages. that each transmission may be controlled to a fine granularity separately. As a result,

service providers will be able to They might provide customized unicast services, as well as on-demand services. For their customers, they provide demand and time-shifted watching options. which are not bound by service costs

This unicast transmission constraint is circumvented. via use of multicast/broadcast transmissions in comparison to unicast delivery to each receiver, simultaneous multicast delivery to multiple receivers lowers the overall cost Bandwidth is necessary.



IP Multicast transmission allows data to be delivered from a sender to several recipients while minimizing the amount of identical copies sent, unlike IP Unicast. The receivers are all assigned to the same multicast organization . A sender sends an IP packet to receipt. The networks forwards a copy of the data to each group of hosts the use of the multicast deal with. Multicastdoes not require a connection.the sender sends data through UDP to several receivers. (userdatagram Protocol) is a protocol for replacing information among customers. not like TCP, the UDP protocol truly makes a pleasant try to send statistics. If a packet is lost due to a transmission fault, it's discarded instead of being applied inside the server, the IP Multicast protocol is applied in the routers of a network. The routers of a network make a reproduction of the multicast packet for each destination recipient automatically. As a result, the number of extra copies sent to any given subnet is limited, thus IP Multicast is far more efficient than IP Unicast, needing far less server bandwidth.

Two techniques are used in IP Multicast routing algorithms and protocols:

multicast organization participants are densely spread at some stage in the network, in step with the dense-mode routing algorithm. To set up and hold the spanning tree, it is based on periodic flooding of the network with multicast traffic.

The multicast organization members are broadly dispersed for the duration of the network, according to the sparse-mode routing algorithm. Flooding would consume network capacity in this situation, resulting in major performance issues. As a result, it employs more selective strategies for establishing and maintaining multicast trees.

The Internet's Multimedia Broadcasting:

Access to a cutting-edge broadcast channel will expand their audience's reach. In fact, 60% of online radio listeners reside within the radio's emission area. As a result, employees may now listen to personal radios at work, a trend that is gaining traction in the United States. The cost of advertising will rise as the size of the audience and the length of time they spend listening grows. The addition of actual-time broadcasting radio to the internet webpage will beautify the quantity of vi sitors to the webpage and, as a end result, the advertising revenue generated by way of it.

The media delivery these broadcasters established can be characterized by four key attributes. It was unimodal because broadcasters dealt with a single (terrestrial) transmission path. It was proprietary, as they operated their own hardware-centric infrastructure for the sole purpose of transmitting their channels. Even though the BBC and ITV would eventually share some transmission stations, each operated their own network (Sendall, 1983: 2). Media delivery was vertically integrated because it was a

task that broadcasters carried themselves and were equipped to do. The earliest 'time-shifting' technologies were videocassette recorders (VCRs) and later digital video recorders (DVRs). Video on Demand was created by cable companies (VOD). They originally provided 'near VOD' by displaying content carousels with new films starting every few seconds. When bandwidth expanded, genuine VOD was given for 20 minutes. Despite the fact that IP delivery It's been through an evolution, but it's still a game changer. IP transmission is not only comprehensive but also cost-effective. It revolutionized media distribution, shifted business structures, and ushered in a new era of video. In terms of quantity and complexity. the ecosystem significantly outnumbers the broadcasting industry. Ecosystems from the recent past Several variables contribute to the complexity. First, the number of video creators has increased.

Furthermore, VOD is a 24-hour environment in which viewers expect to be able to access or download content at any time and from any place, whether they are at home or on the go. Mobile phones and tablets, as well as gaming consoles and consist of broadcasters' and pay-tv vendors' VOD offerings, social media networks (including instagram or Snapchat), video-sharing apps and structures (inclusive of YouTube, vimeo, dailymotion), and tech and media conglomerates' direct-to-

patron (DTC) systems (Amazon Video,Apple tv+, Disney+, AT&T's HBO Max), and so forth; Cunningham and Craig, 2019; Rein and Venturini, 2018; Spilker e there is additionally increasingly more rights holders with their personal video-enabled websites, ranging from sports leagues to news groups. live streaming (that is significantly employed through distinguished video -sport focused streaming web sites), transactional VOD, and television channel distribution are just a few of the programs that video-sharing lends itself to (Fontaine et al,2018).ultra-high definition (HD) one zero five-

inch television presentations,are increasingly being ap plied to get entry to facts. in the end, than in the beyond, audiences are extra fractured and dispersed. Transnational tv networks, together with CNN, Discovery, and Disney, as well as enjoyment platforms, must attain humans from all over the global. Even countrywide broadcasters that used to simply have one feed in step with channel all through the analogue era now have many local and sub-nearby feedsto cater to neighborhood tastes and demands (Chalaby, 2019).

II. LITERATURE SURVEY

Various strategies for user-centric multimedia transmission via the Internet have been thoroughly investigated. The energy saving options for battery-powered mobile devices are covered in Hoque et al survey on multimedia streaming . The writers supplied the following: a wide categorization of current solutions, such as as a result of the OSI model's layers and cross-layering adjustments at the layer level, based on energy-efficient traffic schedule without changing the wireless channel the properties of multimedia (encoding parameters, etc.) based on adaptive multimedia and a high transmission rate a method of delivering multimedia that alters its properties content for lower energy usage at the user's end The media transcoding, application layer solutions on SVC The selection of multimedia transmission characteristics is dependent on based on device specifications, network bandwidth, and Heterogeneity of CPUs. The planning and structuring of traffic The optimization of end points is a the heart

of the solutions. Specifically, pure client-centric, proxy-centric, and server-client architectures.

Wireless Networks Use IP Multicasting for Mobile Hosts:

The undertaking of turning in multicast to mobile hosts making use of cellular IP for community routing aid is presented in this study. because many multicast strategies are in efficient whilefaced with common membership or reg ion changes, presenting multicastin an internetwork with mobile hosts is particular.

If we want to present extremely good mobility support to our community, we cannot appoint mobile IP without making great ad justments. mobile IP, especially, does now not scale to the size of the mobile community due to the fact binding updates must be transmitted to each out of doors node that corresponds to a cellular network node. Periodic binding updates would over burden the spine community when you consider that a cellular community will have loads of nodes and every node within the MN can connect with many CNs. As a result, Mobile For IP Multimedia Subsystems, a Dynamic Video Streaming Quality Adaption Method:

A dynamic video streaming fine edition approach for real-time video streaming service in IMS became supplied. They make modifications to the MCMS so that it could reveal heritage visitors and perform service first rate management. The usage of adynamic video fine transfer can assist to decrease bandwidth ut ilization.The simulation consequences reveal that the recommended strategy correctly avoids communit ycongestion and improves QoS. A excessive amount of bandwidth is needed for real-time multimedia services along with video streaming. Further more, the numerous on-going media classes may additionally result in traffic accidents. We present a dynamic video streaming best model technique for IMS in this research. whilst the overall network load is excessive, the video nice will dynamically adapt to the history site visitors, decreasing bandwidth congestion. The simulation results demonstrate that by sing implementing the suggested adjustments.

IP Multicasting in HF Radio Networks:

A multicast records connection protocol for 0.33-era (3G) excessive-frequency (HF) radio networks and assess its abilities to handle IP multicasting of navy. communications in these networks. This study presents and evaluates a proposed Multicast Data Link (MDL) protocol that would supplement NATO STANAG 4538's existing 3G HF protocol suite. The planned MDL protocol offers a 3G alternative to STANAG 5066's popular 2G HF data connection, and it should be interchangeable in terms of military message (STANAG 4406). MDL fully interacts with the connection setup and traffic management aspects of both "robust" and "quick" 3G HF protocols, and allows for a decade-long tradeoff between robustness and speed.

Internet Multicast Video Delivery:

Two architectures for multicast support have been proposed: IP and overlay multicast. For the past 15 years, Internet video distribution has fueled research in multicast routing, quality of service, and the Internet's service paradigm. Multicast distribution offers the ability to distribute a vast quantity of content that is currently unavailable for broadcast transmission. Since the late 1980s, a significant amount of research has been did on IP multicast and qu techniques for IP multicast. Overlay multicast research has gained traction in the last five years,

with the goal of achieving ubiquitous multicast distribution that is efficient and scalable in terms of the number of groups, receivers, and senders. This paper provides a high-level summary of the difficulties.

III. DESIGN AND IMPLEMENTATION

The code largely adheres to the standard design. It does not employ multi-threading and instead uses numerous processes. The procedure of terminating a station is done by killing it.

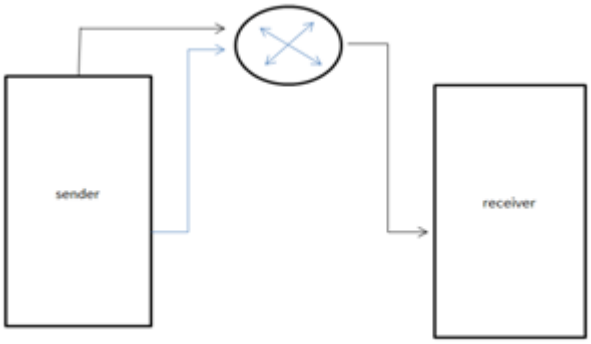
The app provides a menu with choices for pausing, resetting, changing stations, and closing.

The following is the sequence of events:

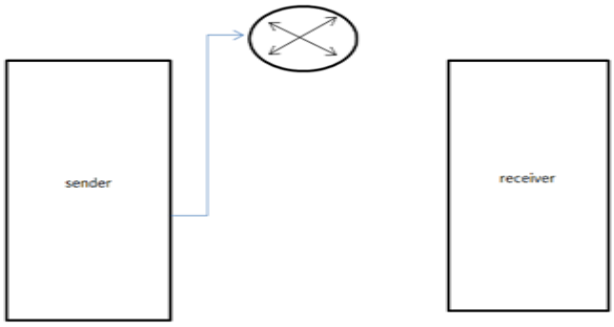
- Client builts a connection with the server;
- Server transmits channel data to the client;
- Client builts a connection with the specified channel;
- Client begins receiving packets in a buffer.

System Design

Phase 1: sender sends TCP packet to receiver with station information.

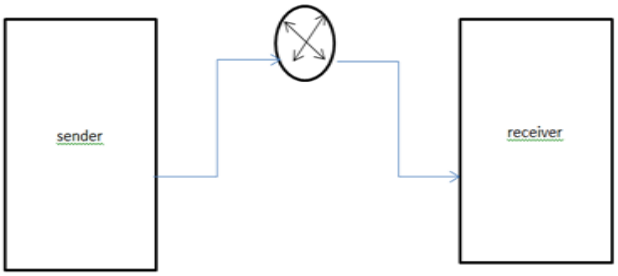


Phase 2 : Receiver selects preferred channel

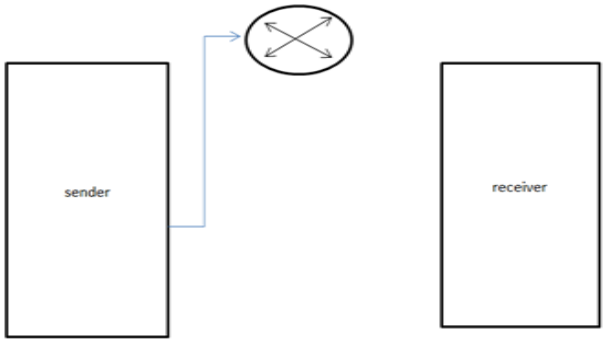


Phase 3: The station music is received by the receiver.

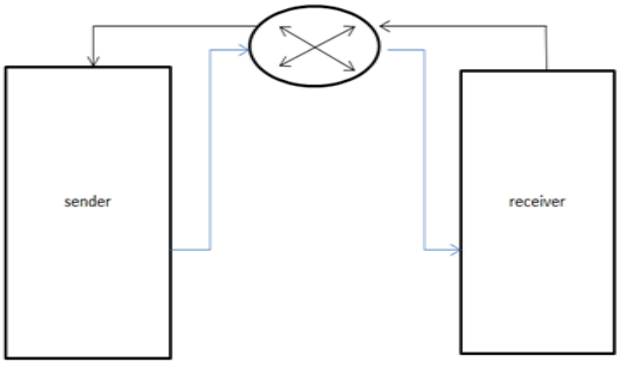
Reset, Pause, Change Station, and Close (Phase 4)



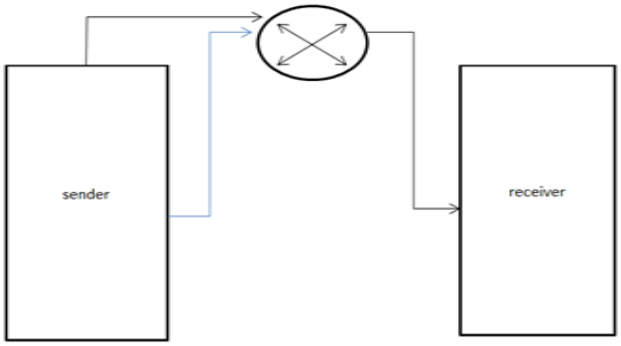
Pause



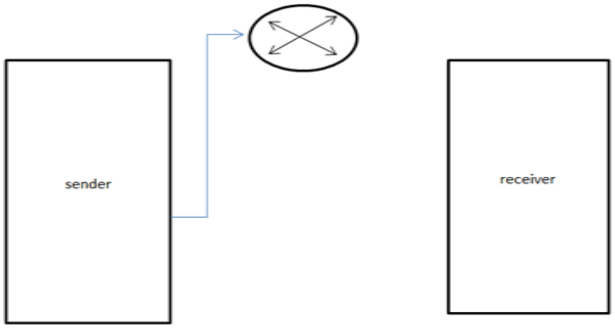
Change Station: A TCP packet is sent by the receiver to request a change in station.



Sender transmits the station list in a TCP packet.



Close station:



IV. CONCLUSION

Ip multicasting is the advanced approach for actual time data broadcasting onthe net.

it enables an efficientcoverage of “massive activities” a the web, without additional investment to the server and network on the internet including radio and television broadcast, real-time broadcasting of essential facts, distance gaining knowledge of, and plenty of others Bandwidth. The variety of connection at the net , protected with the aid of Ip multicast, is nearly unlimited. coupled with green audio and video compression strategies.

V. REFERENCES

[1]. Y. Chawathe, S. McCanne, and E. Brewer. An architecture for Internet content distribution as an infrastructure service,” http://www.cs.berkeley.edu/~yatin/papers/scattercast.ps

[2]. H. Schul zrinneet al., “RTP: A Transport Protocol for Real-Time Applications,” RFC 1889, IETF, Jan. 1996.

[3]. Kai-Di Chang, Chi-Yuan Chen, Shih-Wen Hsu, Han Chieh Chao, and Jiann-Liang Chen, "Advanced Path-Migration Mechanism for Enhancing Signalling Efficiency in IP Multimedia Subsystem," KSII Transactions on Internet and Information Systems, Vol. 6, No . 1, pp. 305-321, January 2012

[4]. H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[5]. Morris, S. Chaigneau, A.: Interactive TV Standards, Focal Press, 2005, ISBN 0-240-80666-2.

[6]. J. Bani-Salameh, Military Multipath Multicast Routing Protocol, PhD Dissertation, New Mexico State University,2005.

[7]. Interactive Streams [Online]. Available: http://www.interactivestreams.com/.

[8]. Charles E. Perkins, ”Mobile Networking Through Mobile IP,” IEEE Internet Computing, Vol.2, pp.58-69 Jan./Feb. 1998.

[9]. S. Banerjee, B. Bhattacharjee, and C. Kommareddy. Scalable application layer multicast. In Proceedings of ACM SIGCOMM 2002, Aug. 2002.

[10]. Lin Xu, Ralf Tönjes, Toni Paila, Wolfgang Hansmann, Frank, Markus Albrecht. “driving to the Internet: Dynamic Radio for IP Services in Vehicular Environments” Proc. of the 25th Annual Conference on LocalComputer Networks, LCN'00, Tampa, FL, November 2000.

[11]. B.Furht, R. West water, and J. Ice, “Multimedia Broadcasting over the Internet: Part I,” ZEEE Multimedia, Project Reports, Vol. 5, No. 4, Fall 1998, pp. 78-82.

[12]. S. Deering, Host Extensions for IP Multicasting, RFC 1112, 1989[Estrin et al. 98].

[13]. D.Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei, Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification, RFC 2362, 1998

[14]. Lin Xu, Toni Paila, Wolfgang Hansmann, Matthias Frank. IPv6 based Infrastructure for Wireless IP in Multi-Radio Environments with Quality of Service Support. Proc. Of the 26th Annual Conference on Local Computer Networks, LCN'01, Tampa, FL, November 2001.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research inScience, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Graphical Password Authentication System

Bhavana V[1], Dr. M Nirmala[2], Manikandakumar M[3], Ms. Soja Rani[3]

[1]PG Student, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Associate Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[3]Senior assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Graphical password authentication provide alternative to traditional alphanumerical password. This method has vast drawback. Since people remember pictures better than words. When any application is provided with user friendly authentication it becomes easy to access and use that application. one of the major reasons behind this method according to psychological studies human mind can easily remember images than alphabets or digits. To address this problem some researchers have developed authentication methods that uses pictures as password. The basic goal of this system is to achieve higher security with simple technique to use by a user and harder to guess by a hacker. Graphical password authentication system is best alternative for text password.There are two types of tactics: recognition-based and recall-based techniques. We look at the benefits and drawbacks of each technique, as well as possible future research avenues in this area. We also try to answer for question "What are the key design and implementation challenges for graphical passwords?" This poll will help datasecurity researchers and practitioners decide whether or not to use text-based authentication systems.

Keywords- Authentication, graphical passwords, usability, security, images, recall based technique, recognition based technique

I. INTRODUCTION

For security reasons, every web application that wants to access the network requires user authentication. There are a variety of authentication mechanisms available to ensure security, and alphanumeric passwords which has high security, sometimes known as text-based passwords, are one of them. Since ancient times, secret code or data has been employed to obscure and secure information. Knowledge-based authentication, which incorporates both alphanumeric and graphical passwords, is available in the majority of web applications. A knowledge-based authentication system's most significant usability goal is to help users choose higher-security passwords with greater password space. guess images. The task of choosing a weak password that is easy to guess by attacker is more time consuming in the suggested

system, discouraging users from making these decisions.

Text-based passwords, on the other hand, are easily guessable and accessible via brute force attacks, dictionary attacks and other methods. To give better protection, an alternative method for text-based alphanumeric passwords has been developed. It has been designed a graphical password authentication system. The user must choose a password from a sequence of photos in a certain order in this authentication scheme. The images which should be chosen can be of any type, including images of nature, places, and so on. the user is unable to identify the image during logging, the previously clicked image is erroneous and that the user should resume the password entering,whereas explicit feedback is offered after the last click point. As a result, a graphical password can avoid many of these attackers.

RESISTANT TO POPULAR ATTACKS

1. Brute force: When the user reaches the maximum number of tries, he or she will be automatically notified. Furthermore, further authentication via the generic Webpage is disabled for such a registered user; instead, they must use the link provided by the company in the email notification. This also informs the legitimate user of the adversary.
2. Shoulder surfing: Shoulder surfing is a social engineering technique that involves looking over the victim's shoulder to collect information about the personal identification numbers (PINs), passwords, and other confidential data. We use a system similar to the Mobile pattern system. When users draw the pattern, that is invisible on the screen. This makes it extremely difficult for the opponent to see the pictures on the grid that user clicks.
3. Spyware: Graphical password systems are more resistant to spyware than traditional passwords. Keyloggers secretly capture and transfer keystrokes, and if the spyware intends to track mouse movements, it can, but the adversary can't determine which mouse event is the graphical password. The timeline vs mouse-event graph is too hard to understand.
4. Hidden camera: A camera will be placed in front of the user to identify a face during authentication, i.e., the number of pixels occupied by the face should be 80-90 percent of the total pixels in the current frame, and if this condition is not met, the screen will not display the graphical password. It prompts the user to cover the screen properly. However, this will be an expensive operation.
5. Phishing: Since this adversary is led to believe that password is a collection of images, it is impossible to create a bogus page because the adversary believes he does not know the images. Furthermore, we limit the user to try once and advise the user to use a fake password each time so that the server sends an email with a URL so that he can sign in via the genuine login page and the malicious user cannot send the Link to users from a genuine server. However, if the adversary is aware of the technique, this attack may still be useable.

II. LITERATURE SURVEY

A. Recognition Based Techniques

Dhamija and Perrig [1] proposed a Hash Visualization-based graphical authentication scheme. In their system, the person is allowed to select a specific number of images from a set of different images generated by a program (figure 1). Later, in order to be authenticated, the user is required to recognise the pre-selected images. The results showed that 90 percent of all members successfully authenticated using this approach, while only 70 percent successfully authenticated using PINS and text-based.

Figure 1. Dhamija and Perrig

Sobrado and Birget [2] created a graphical password technique that addresses the shoulder-surfing issue. The system displays a number of pass-objects (pre-selected by the user) amongst many various objects in the first scheme.To authenticated, a person must recognise pass-objects and click inside the curve formed by all pass objects (figure 2). To make the password difficult to guess, Sobrado and Birget proposed using 1,000 objects, where it makes the screen very crowded and the objects almost indistinguishable; however, using fewer objects may result in a smaller password space, given that the resulting edge can be large. The authors also advise repeating the process will limit reduce the possibility of logging by rotation and clicking randomly. This algorithm makes login system slow.

Hong et al. [3] subsequently applied this method to allow users to assign personal codes to pass-object variants. Nonetheless, this method forces the individual to memorize a large number of text - based strings and thus go because of the numerous disadvantages of text-based passwords

Figure 3. Another shoulder sur ng resistant scheme developed by Hong, et al. [3].

Weinshall and Kirkpatrick [4] developed hundreds of new authentication schemes, including picture recognition, pseudo word recognition, and object recognition, and conducted hundreds of new user studies. A user is trained to recognise a massive set of pictures (100 – 200 images) chosen from a data set of 20,000 images in the picture recognition study. Users in their study were able to recognise more than 90% of the pictures in the training set after one to three months. According to the findings of this study, images are the most powerful of the three schemes tested. Pseudo codes also can be used, but suitable training is required.

“passface" is a scheme created with the help of Real User Corporation [5]. The basic concept is as follows. The individual will be asked to choose four face images of human from a set of four their future password, they will use the face database. During the authentication process this point, the individual notices a grid of 9 faces, each of which is made up of one face until now, the person and 8 decoys have chosen the person's face as shown in (Fig. 4) The

individual recognizes and clicks anywhere the identified face This procedure is repeated numerous times. the user successfully logs in, he or she is authenticated and identifies each of the four faces The strategy is based on the is less difficult to assume that humans can remember human faces.

Figure 4. Passfaces

Jansen et al.[6] proposed a mobile device graphical password method. During the enrollment stage, a person chooses a theme (for example, sea, cat, etc.) that consists of thumbnail images and then registers a sequence of pictures as a password (figure 5). During authentication, the user must enter the registered pictures in the correct order. the disadvantage of this method is that because the number of thumbnail images is limited to 30, the password space is limited. Each thumbnail image has a numerical value assigned to it, and the sequence of choice generates a numerical password. The end result revealed that the length of a photo sequence used to be usually smaller than the length of a video sequence. length of textual password, Two solutions have been proposed to address this issue. Images can be combined to form a new alphabet element, as shown below. As a result, the photo alphabet's size has grown.

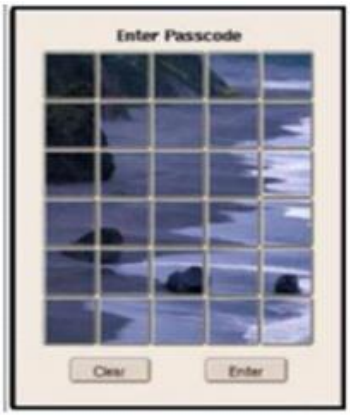


Figure 5. A graphical password scheme proposedby Jansen, et al.

B. RECALL BASED TECHNIQUE

In this section, we will look at two kinds of picture password techniques: repeating a selection and reproducing a drawing. Jermyn et al. [7] proposed a technique called "Draw - a - secret (DAS)" that allows the user to draw their own password (figure 6). A user is asked to draw a simple image on a 2D grid. The positions of the grids occupied by the photo are saved in the drawing's order. During authentication, the user is asked to redo the image. If the user's drawing touches the identical grids in the same order, the user is authenticated. According to Jermyn et al., given reasonable-length passwords in a 5 X 5 grid, the full DAS password space is larger than the full textual content password space.

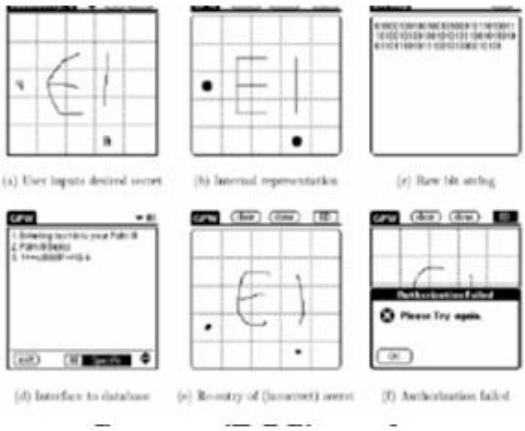


Figure 6. Draw-a-Secret (DAS) technique proposed by Jermyn, et al.

van Oorschot and Thorpe [8] investigated the memory capacity. Jermyn et al. defined the password space of the graphical password scheme. They pioneered the concept of graphical dictionaries, as well as researching the possibility of a brute-force attack the application of such dictionaries They specified a length. parameter for the DAS type of graphical passwords and displayed that passwords of length DAS is eight or greater are displayed on a 5 x 5 grid may also be less vulnerable to dictionary attacks than passwords in text They also demonstrated the space It is significantly easier to replicate symmetric graphical passwords. less than the entire DAS password space Because people recall It is more common to rank symmetric images higher than asymmetric images. It is assumed that a large proportion of users will select replicate. passwords that are symmetric. If this is the case, the DAS scheme's protection may be significantly lower than previously thought. This issue can be resolved by using longer passwords. Thorpe and van Oorschot demonstrated the size of the space of recto symmetric passwords of size approximately L + 5 exceeds the size of the full password area for corresponding length L = 14 on a 5 x 5 grid.

van Oorschot and Thorpe [9] investigated the impact of stroke-count and password length as a complexity. The DAS scheme's property. Their ND out about demonstrated that Thestroke-count has the greatest influence on the DAS. password area DAS password area measurement decreases significantly with lesser strokes for the same amount of time length of the password The DAS password length is also important. The impact is significant, but not as strong as the stroke count. van Oorschot and Thorpe worked together to improve security. A "Grid Selection" method was anticipated. The decision grid is initially a large, fine-grained grid from which the user selects a drawing grid to zoom in on triangle region, and a password entry field (figure 7). This would significantly increase the password space of the DAS.

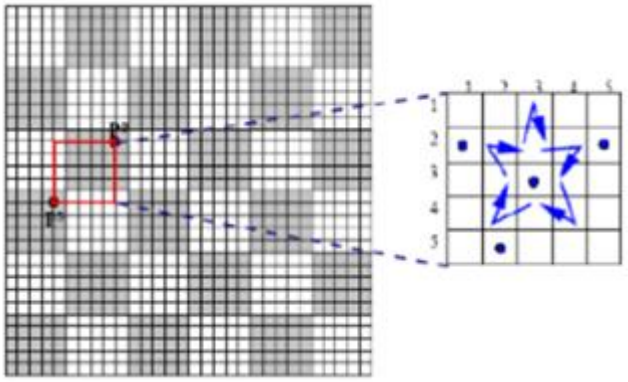


Figure 7. Grid selection

Goldberg et al. [10] conducted an user study in and they used a method known as "Passdoodle." This is a graphical password made up of handwritten designs or text drawn with a stylus onto a sensitive touch screen. According to their findings, users can remember entire doodle pictures as exactly as alphanumeric passwords. The person studies also revealed that humans are much less likely to remember the sequence in which they doodled a DAS password. However, user study was conducted using a paper prototype rather than computer programs, with verification performed by a human rather than a computer, the accuracy with this study is still unknown.

C. REPEAT A SEQUENCE OF ACTIONS

Blonder [11] created a graphical password scheme in which the user creates a password by clicking on various locations on an image. During authentication, the individual must click on the approximate areas of those locations. Because the image can assist users in recalling their passwords, this approach is regarded as more convenient than unaided recall (as with text-based password).

Passlogix [12] has created a graphical password device based primarily on this concept. In their implementation (figure 9), users must authenticate

themselves by clicking on a number of objects in the image in the correct order. In order to determine whether item is clicked or not with the mouse, invisible boundaries are defined for each item. A similar method has been developed by sfr. Microsoft was said to have developed a comparable graphical password approach in which users are asked to click on preselected areas of an image in a specific sequence. However, key aspects of this method are not available. Wiedenbeck et al.'s [13] "PassPoint" system expanded Blonder's concept by removing the predefined boundaries and allowing the use of arbitrary images. As a result, instead of some pre-defined areas, a person can create a password by clicking on any area on a picture. A tolerance is calculated around each selected pixel. To be authenticated, the person must click within the tolerance of their chosen pixels and also in the correct sequence (figure 10). This method is entirely based on the discretization method proposed by Birget et al. Because any photo can be used, and because a picture can contain hundreds or even thousands of memorable points, the password space is quite large. Wiedenbeck et al. conducted a study in which one group of participants was asked to use an alphanumeric password, while another group was asked to use a graphical password. The end result demonstrated that graphical passwords required fewer attempts by the user than alphanumerical passwords. However, graphical password users had much more difficulty learning the password and took longer to enter their passwords than alphanumerical password users.

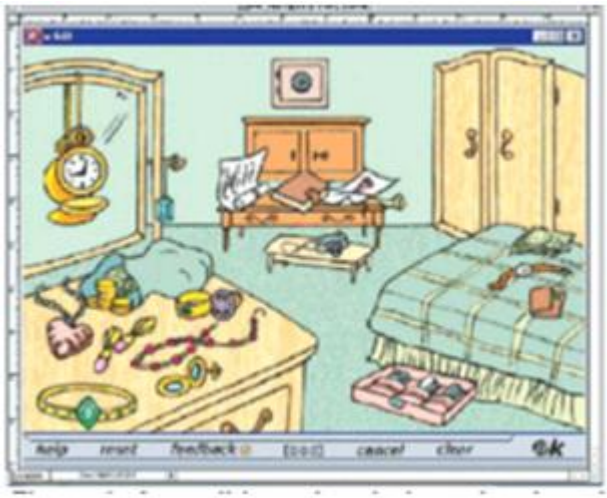


Figure 8. A recall-based technique

III. DISCUSSION

Is it as safe to use a graphical password as a text-based password?There has been very little research done to learn more about the problem of cracking graphical passwords. Because graphical passwords are not widely used in practice, there is no data on actual cases of graphical password cracking. In this section, we look at some of the possible techniques for cracking graphical passwords and compare them to text-based passwords.

GUESSING

Unfortunately, it it seems that graphical passwords are regularly predictable, which is a serious issue with text-based passwords. Studies on the Passface approach, for example, have revealed that people frequently choose vulnerable and predictable graphical passwords. Nali and Thorpe's research discovered comparable predictability among graphical passwords generated using the DAS method. More research is needed to understand the nature of graphical passwords generated by real-world users.

SOCIAL ENGINEERING

When compared to text-based passwords, graphical passwords are less useful for a person to give to another person. For example, it is extremely difficult

to give out graphical passwords over the phone. It would take more time to set up a phishing web page to obtain graphical passwords. Overall, we believe it is extremely difficult to crack graphical passwords using standard attack methods such as brute force search, spyware and dictionary attacks. There is a need for more in-depth research into possible attack strategies against graphical passwords.

DICTIONARY ATTACKS

Because recognition-based graphical passwords require mouse input rather than keyboard input, dictionary attacks against such a type of graphical password will be impractical. A dictionary attack is possible for some recall-based graphical passwords, but an automated dictionary attack are much more complex than a text-based dictionary attack. More research in this area is required. In general, we believe that graphical passwords are less vulnerable to dictionary attacks than text-based password.

IV. PROBEM DEFINITION

To address the shortcomings of existing text-based password systems and to make the systems more secure from hacking and predictable, we are developing a new system. a new generation of passwords based on images that cannot be predicted or hacked easily Solution to the above-mentioned system limitation. In order to compensate for the drawbacks of the existing text-based password systems and to create new ones the systems are more secure against hacking and predictability We're working on a new generation of passwords based on images that can't be predicted or hacked. They can't be written in any way, so there's no way to steal them. Security researchers and practitioners have conducted research on how to protect systems, individual users, and digital assets. The password issue stems primarily from limitations in human long-term memory (LTM). Once a password has been selected and memorized, the users should be able to recall it in order to log in. However, people frequently forget their passwords. Graphical passwords usher in a completely new type of authentication. The most common type of authentication used today is alphanumeric text, and this type of authentication has been shown to be vulnerable to a variety of attacks. Because pictures are easier to remember than text, graphical passwords tend to improve security while also making it easier for users to use and access data.

V. IMPLEMENTATION

Through the use of click points on the image password, a secure graphical authentication system is provided that protects users from becoming victims of attacks when entering passwords in public. The next image is displayed in this scheme based on the correct location of the previous click-point. The user will be authenticated after correctly clicking on the password, or vice versa. After three incorrect inputs, the user will be blocked. The user can be unblocked by the administrator, or the user must go through three layers of security, which include an OTP, a security question, and an image password. This system is more user-friendly and simple to use. Because users use a dynamic click point to indicate the position of their password, this provides better security against dictionary attacks and brute force attacks. It has a wide range of applications in web-based applications. First, there is the graphical login page, which appears at the start of the application. On this page, there is a panel with 25 images. In this manner, a total of eight panels with 25 images each are used. To set the password, the user must first choose an image from the panel. When you select this image, the images are shuffled. The user must recall the sequence of image selections that he or she chose as their password. The

user clicks the login button after selecting 8 images from the 8 panels. After clicking the login button, a single image panel will appear, and the user must choose any portion of the image as a password. As a result, the user's graphical password is saved, which is made up of 8 random images and one single part of the entire image. After clicking OK, the user's password is saved and the banking application is launched, allowing the user to perform in it or work on tasks as needed.

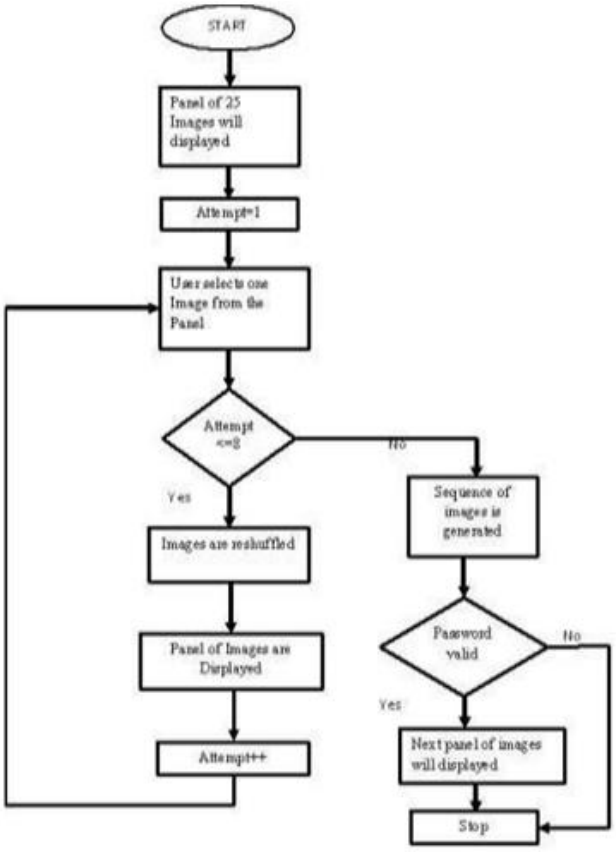


Figure 9: Flowchart

VI. CONCLUSION

Over the last decade, there has been an increase in the use of the graphical passwords, as an alternative to the conventional text-based passwords We conducted a comprehensive survey of current graphical password techniques in this paper. Modern graphical password methods can be two types of recognition systems: recognition-based and recall based technique Despite the fact that the primary argument for graphical passwords is that people have a better memory for graphical passwords. Current user studies show that text-based passwords are more secure than text-based passwords are limited, and there is no longer compelling evidence to guide this point of contention According to our preliminary findings, it is extremely difficult to crack graphical passwords using standard attack techniques such as dictionary attack, brute force search, or spyware. However, due to the limited implementation of graphical password systems, the vulnerabilities of graphical passwords are not fully understood. Overall, modern graphical password techniques are still in their infancy. Much more research and user testing are required for graphical password techniques to mature and become useful.

VII. REFERENCES

[1]. R. Dhamija and A. Perrig, "Deja Vu: A User Study Using Images for Authentication," in Proceedings of 9th USENIX Security Symposium, 2000.

[2]. L. Sobrado and J.-C. Birget, "Graphical passwords," The Rutgers Scholar, An Electronic Bulletin for Undergraduate Research, vol. 4, 2002.

[3]. D. Weinshall and S. Kirkpatrick, "Passwords You'll Never Forget, but Can't Recall," in Proceedings of Conference on Human Factors in Computing Systems (CHI). Vienna, Austria: ACM, 2004, pp. 1399-1402.

[4]. D. Hong, S. Man, B. Hawes, and M. Mathews, "A password scheme strongly resistant to spyware," in Proceedings of International conference on security and management. Las Vergas, NV, 2004.

[5]. Real User, "www.realuser.com," last accessed in June 2005.

[6]. W. Jansen, S. Gavrila, V. Korolev, R. Ayers, and R. Swanstrom, "Picture Password: A Visual Login Technique for Mobile Devices," National Institute of Standards and Technology Interagency Report NISTIR 7030, 2003.

[7]. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin, "The Design and Analysis of Graphical.

[8]. J. Thor pe and P. C. v. Oorschot, "Graphical Dictionaries and the Memorable Space of Graphical Passwords," in Proceedings of the 13th USENIX Security Symposium. San Deigo, USA: USENIX, 2004.

[9]. J. Goldberg, J. Hagman, and V. Sazawal, "Doodling Our Way to Better Authentication," presented at Proceedings of Human Factors in Computing Systems (CHI), Minneapolis, Minnesota, USA., 2002.

[10]. J. Thorpe and P. C. v. Oorschot, "Towards Secure Design Choices for Implementing Graphical Passwords," in Proceedings of the 20th Annual Computer Security Applications Conference. Tucson, Arizona, 2004.

[11]. G. E. Blonder, "Graphical passwords," in Lucent Technologies, Inc., Murray Hill, NJ, U. S. Patent, Ed. United States, 1996.

[12]. Passlogix, "www.passlogix.com," last accessed in Jun 2005.

[13]. S. Wiedenbeck, J. Waters, J. C. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: Basic results," in Human Computer Interaction International (HCII 2005). Las Vegas, NV, 2005.

[14]. S. Wiedenbeck, J. Waters, J. C. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: Basic results," in HumanComputer Interaction International (HCII 2005). Las Vegas, NV, 2005.

[15]. S. Wiedenbeck, J. Waters, J. C. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: Effect s of tolerance and image choice," in Symposium on Usable Privacy and Security (SOUPS). Carnegie Mellon University, Pittsburgh, 2005.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Network Traffic Analyzer - Case Study

Aishwarya S, Dr. Senthil Kumar, Manikandakumar M, Soja Rani
Department of Computer Science and Engineering New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Analysis and prediction of network traffic has numerous applications and has recently attracted a significant number of studies. Experiments of various types are carried out and summed up in order to determine various problems in existing computer network applications. The analysis and prediction of network traffic is a proactive approach to ensuring secure, reliable, and qualitative network communication. Various techniques for evaluating traffic on the network are developed and tested, ranging from neural network- based techniques to data mining techniques. Similarly, various linear and non-linear models for network traffic prediction are proposed. Several intriguing combinations of network analysis and prediction techniques are used to achieve efficient and effective results. This paper aims to provide an overview network analysis and traffic forecasting techniques.

I. INTRODUCTION

Network traffic prediction and analysis represents a proactive rather than a selective approach, in which the network is surveilled to ensure that no data breaches occur within the network. The network traffic analysis phase is critical for designing efficient preventive congestion control strategies and recognising normal and malicious packets.These schemes seek to avoid network congestion by allocating network resources based on expected traffic. The predictability of network traffic has substantial benefits in many regions, including dynamic bandwidth allocation, network security, network planning, and predictive congestion control, among many others.We can split predictions into two categories: long-term predictions and short-term predictions. The process of capturing, filtering, and analysing network traffic is known as network traffic analysis.

Analyzing network traffic can reveal vital information about network device activities, the type of data shared between them, bandwidth utilisation for such transfers, and the detection of anomalies.Long queues forecasting provides detailed forecasting of traffic models to evaluate future capacity requirements, letting for more detailed planning and better decisions. Dynamic allocation of resources is linked to short-term prediction (milliseconds to minutes). It can be used to improve Quality of Service (QoS) mechanisms, as well as to control congestion and manage resources more effectively. It can be used for packet routing. This study gives an overview of several techniques for network traffic analysis and forecasting that have been proposed, used, and performed.

II. NETWORK TRAFFIC ANALYSIS

Network traffic analysis (NTA) is a technique for monitoring the network availability and activity in order to detect anomalies like security and operational issues. NTA is widely used in the following cases:

- Maintaining a real-time and historical record of what's going on in your network
- Detecting malware, such as ransomware activity
- Detecting vulnerable protocols and cyphers
- Troubleshooting a Slow Network
- Improving internal visibility and removing blind spots

Implementing a solution that can continuously monitor network traffic provides you with the insight you need to optimise network performance, reduce your attack surface, improve security, and better manage your resources. Knowing how to monitor network traffic, on the other hand, is insufficient. It's also important to think about the data sources for your network monitoring tool, two common are data flow and packet data. Network traffic analysis is critical for monitoring network availability and activity in order to detect anomalies, improve productivity, and detect malicious. Network traffic is a critical component of comprehensive visibility and security analysis which allows detection and destruction of threats.

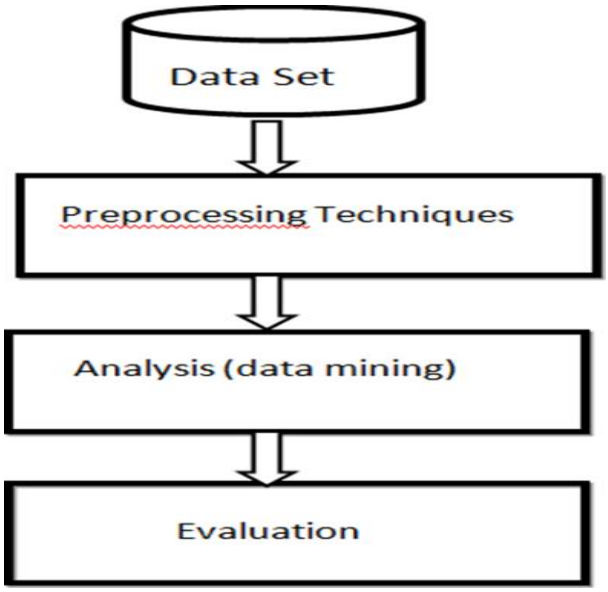


Fig1: Generic Structure of Network Traffic Analysis

2.1. Data sets

1. Data set from DARPA

KDD cup data has been the most widely used for evaluating network traffic analysis with respect to intrusion detection This information is given by Stolfo and colleagues[3]. It is constructed based on the data. DARPA IDS evaluation programme captured .The KDD cup data set consists of approximately 4,900,000 training instances and 41 features. Furthermore, the test data set contains 300,000 occurrences The KDD cup has 24 attacks are being trained and tested, and 14 new types have been added reported in testing results

2. Dataset NSL-KDD

The NSL-KDD data set is a revised version of the KDD cup data set[4]. The NSL-KDD data set does not include duplicate records in testing data or redundant instances in training data. As a result, the classifier improves in accuracy. The NSL-KDD data set is open to researchers and is an enhanced version of the original KDD cup data set.

i. CAIDA data set:DoS attacks were included in the data set[5].

ii. Data set from Waikato:It has a storage area for internet traffic[6].
iii. Dataset from Berkeley Lab:It contains an archive of internet traffic[7].

2.2 Preprocessing Techniques

Pre - processing is a vital step in converting real- world data into such an understandable format. Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format. Certainly, real-world data are frequently incomplete and noisy in specific behaviour. In other words, most of the data we want to analyse from the real world using data mining techniques is insufficient and inconsistent (containing errors, outlier values). As a result, before applying data mining techniques, preprocessing methods are required to improve the quality of the data, thereby assisting in improving the accuracy and efficiency of the resulting data mining task. But since network traffic patterns vary in format and dimensionality, pre - processing techniques are critical and important in network traffic analysis.

1. Data mining

Data mining entails using refined data analysis tools to discover previously unknown, valid patterns and relationships in massive data sets. These tools can include statistical models, machine learning techniques, and mathematical algorithms like neural networks and decision trees. As a result, data mining includes both analysis and prediction. Several major data mining techniques, such as association, classification, clustering, prediction, sequential patterns, and regression, have been developed and used in recent data mining . Data mining, also known as knowledge discovery in databases, is the process of discovering interesting and useful patterns and relationships in large amounts of data in computer science.

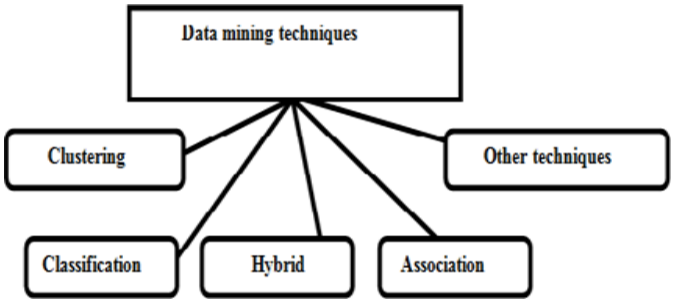


Fig 2. Data Mining Techniques

2. Clustering technique

Clustering is the process of categorising data into groups based on certain characteristics.Clustering divides data into groups of objects that are similar.Every group, known as a cluster, is made up of members who are very similar to one another, and members from different clusters differ from one another. Clustering methods are used to group network data into groups for network traffic analysis.We discovered that researchers use a variety of clustering data mining algorithms. M. Vijayakumar et al. [8] proposed a data mining algorithm for data stream network traffic based on hierarchical clustering. Initially, they used K-means partitioning.Algorithm for forming clusters from network traffic data streams.A suitable cluster is identified and further subdivided.by employing hierarchical techniques to obtain additional informationclusters. This bisection K-Mean, according to the authors,The clustering result outperforms the pure hierarchical result.

2 .3. Classification technique

This method is used to obtain critical and relevant information about data and metadata. This data mining technique aids in the classification of data into various categories. The level of user interaction involved in the data mining procedure, can also be considered in the classification. To detect intrusion, Ghanshyam Prasad Dubey et al. [9] proposed RST and incremental SVM approaches.The authors tested their hypothesis using

the KDD cup dataset. Theselection of significant network traffic attributes.The data set is completed using the RST method, and then SVM approach is used for training and testing.

Hybrid Technique

The hybrid models are a combination of two or more approaches for analysing network traffic.We present several hybrid model techniques being investigated by researchers for network traffic analysis. By combining C4.5 decision tree and Support Vector Machine (SVM) approaches, JashanKoshal et al. [8] proposed a hybrid model for developing an intrusion detection system.

Association rules Data Mining Techniques

Each attribute/value pair is treated as an item by the association rule. An item set is a collection of items in a single network request. Association rules are used to identify patterns or relationships among data base attributes. Association rules are if-then statements that help to demonstrate the likelihood of interactions between data items in large data sets in various types of databases.We examine various association rules algorithms used by researchers for network traffic analysis. To detect normal and malicious packets in network traffic, Zulaiha Ali Othman et al. [9] proposed Fuzzy Apriori, FP-growth, and Apriori association rule data mining. The authors used the Wireshark tool to collect data from the University of Kebangsaan Malaysia (UKM) at various time intervals.

III. NETWORK TRAFFIC PREDICTION

Network traffic prediction is an important issue that has recently piqued the interest of the computer network community. Network traffic prediction is a common issue that can be usedfor network monitoring, network security, avoiding congestion, and increasing network speed.

For network traffic prediction, researchers employa variety of techniques. These techniques are classified into four broadcategories: linear time series models, nonlinear time series models, hybrid models, and decompose models.

Fig 3 depicts the four categories into which various network traffic prediction techniques are classified.

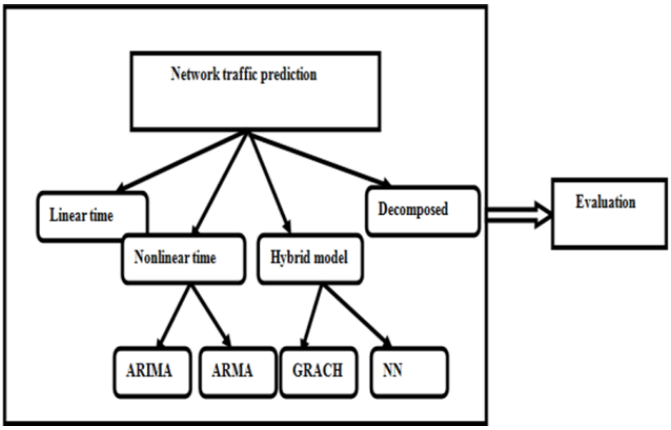


Fig.3 Network Traffic Prediction Technique

1. Technique of Linear Time Series

Linear Time Series Techniques are time series co - variance structures. There are two popular types of linear time series models: auto regressive (AR) and moving average (MA), which can be combined to form auto regressive moving average models. Linear time series are a well-known technique for predicting network traffic. We present several linear time series techniques used by researchers for network traffic prediction.N. K. Hong et al. [10] developed the ARMA time series model to predict large file transfers in network traffic.

2 Techniques for nonlinear time series

Nonlinear dynamic equations cause nonlinear time series to emerge. This model predicts network traffic using various techniques such as neural networks and fuzzy logic.We present a variety of non-linear time series techniques used by researchers for internet traffic prediction. To capture the bursty nature of Internet traffic, Nikkei C.anand, Caterinal et al. [11] proposed a nonlinear time series Generalized Auto

Regressive Conditional Heteroskedasticity (GARCH) model.

3. Hybrid model techniques

A hybrid model is a mix of two or more models. In terms of predicting network traffic, the hybrid model is extremely accurate. The combination of linear and nonlinear functions is a hybrid model. It produces positive results in network traffic prediction and analysis. We demonstrate a variety of hybrid model techniques that are found by Researchers who look into network traffic prediction. S.M.Aqil Burney et al. [12] used an HTTP request series to apply the conditional mean and conditional variance model (ARIMA with GARCH).

Decomposed Model

The time series can be divided into four parts. Each component is described in detail below.

Component of Trend

The trend in time series data is long-term propensity, increase and decrease. The structural variations of low frequency time series are represented by the trend component.

Cyclical component

The cyclical pattern exhibits rises and falls over an unspecified time period.

Seasonal element

Seasonal component refers to variations in time series data caused by seasonal factors. The seasonal data has a stable variation within the time series.

Unusual component

The irregular component is the residual time series after removing the trend and seasonal components. Cheng GuangJaian et al [13] proposed a non-linear decomposed model for predicting long-range network traffic by decomposing time series into trend, period, mutation, and random components.

Evaluation metrics

Many different metrics are used in network traffic prediction techniques to investigate the quality of time series forecasting. The detection rate, false positive rate, accuracy, and time cost metrics are used to assess classifier performance for various data sets. Each metric is defined below.

a. Mean Absolute Error (MAE) It is calculated by dividing the absolute value plus forecasting error by the number of forecasts.
b. (Mean Square Error) (MSE) It is calculated by dividing the number of forecasting values by the number of actual values.
c. (Root Mean Square Error) (RMSE) It is defined as the ratio of observed plus actual values divided by the number of forecasting values. The Root Mean Square Error (RMSE) (also known as the root mean square deviation, RMSD) is a measurement of the difference between values predicted by a model and actual values observed in the environment. This means that the RMSE is most useful when large errors are mostly negative.
d. (NRMSE) Normalized root mean square error. It is the ratio of the RMSE result divided by the number of maximum observed values plus the number of Minimum observed values of forecasting values. (e)(MPE) Mean Percentage Error
e. It is the ratio of actual values plus forecasted values divided by forecasted values at a given point in time. The mean absolute percentage error frequently affects the MPE statistic (MAPE).
f. (f) (MAPE) Mean Absolute Percentage Error It is the ratio of the original values plus the forecasting values divided by the original values, which are then divided by the forecasting values. It is one of the useful evaluation metrics that is unaffected by the forecasting series numbers.

IV. PROBLEM DEFINITION

The goal of this project is to examine the captured packets in order to determine which users were downloading software from blacklisted/illegal websites. The location of the user is determined by analyzing the source and destination IP addresses. We track the user's packet traversal to determine which users are downloading software/data from blacklisted sites. We also created a database to store the geographical locations as well as all of the other information about the packet for future use.

V. RESULTS AND IMPLEMENTATION

1. Proposed System

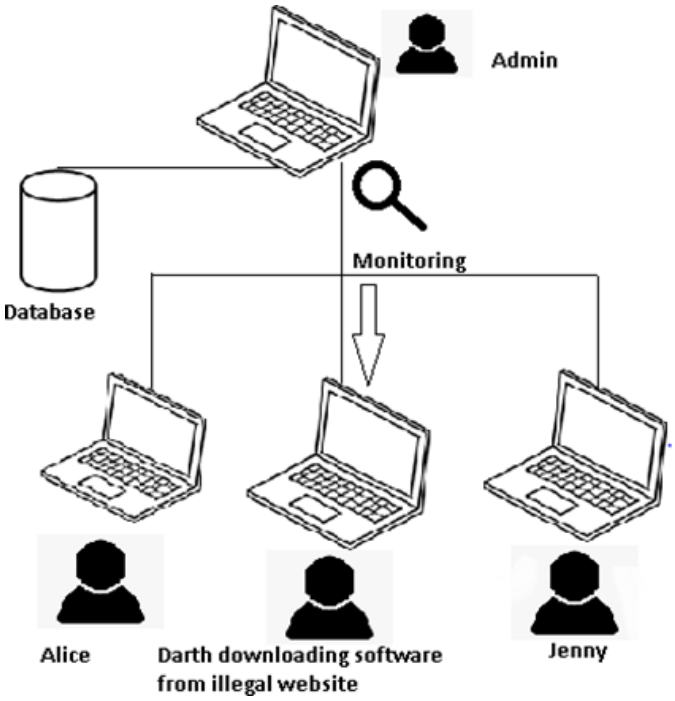


Fig 4 : General view of the Network Traffic Analyzer

This diagram depicts how the application works in general. The administrator continuously monitors the network's users' various activities. If a user engages in any illegal activity, such as downloading from illegal or blacklisted websites, it is detected at the admin level. The administrator then correlates the IP address to the physical location of the packet's destination using PyGeoIP. This is accomplished by querying the database with a specific IP address. The database returns a record with the city, region name, postal code, country name, latitude and longitude plotted on Google Earth. The URL, destination IP, timestamp, latitude, longitude, and so on are all stored in a database.

2. Implementation

Initially, wireshark is used to capture packets, and Ettercap is used to capture live packets.Dpkt is a Python module that is used as a packet analysis tool. It examines each individual packet and the protocol layer. It displays the IP address of the user who downloaded the application from an illegal or blacklisted website. PyGeoIP , which queries the GeoLiteCity database, can be used to determine the geographical location of the packet.In KML format, Google Maps API provides a geographical display of the destination of the packet being analysed. The database stores all information about the packet, including the source and destination IP addresses, timestamps, URLs, geographical latitude, longitude, and plotted location. For security reasons, all information stored in the database is encrypted.

VI. FUTURE WORK SUGGESTIONS

The following methods can be used to broaden the scope of our project:

1. Adding more features to the existing ones
2. Introducing new features

Adding more features to the existing ones:

We are currently using the GeoLiteCity database by MaxMind.com in its free version. GeoIP database, a commercially complete version of the IP geolocation database, can be used in its place. This database provides more precise geographical location information for blacklisted websites and their users. On the local machine, the application has been deployed and tested. This can be deployed on cloud infrastructure like Amazon AWS and accessed from multiple systems.

Introducing new features:

A pcap file is a repository that contains a large amount of information about computer network communication between nodes. From the packets, confidential information such as user credentials (username and password) can be extracted. The packets in the pcap file can be used to analyse network attack behaviour. When a large number of packets originate from a single source IP address, it is possible to detect a Denial of Service (DOS) attack.

VII. CONCLUSION

In this project, to keep a trace of users who are accessing malicious websites, we made and implemented a Network Traffic Analyzer and parsed network traffic using an application code script, extracted IP addresses and other useful informations of users who are trying to access malicious websites and also the IP address of malicious websites . We also obtained the geographical locations . We developed this application with the help of a Python script, which uses GeoLiteCity database, containing all the information regarding the location all across the globe.

VIII. REFERENCES

[1]. M. Vijayakumar, R.M.S.Parvathi, Concept Mining of High Volume Data Streams in Network Traffic Using Hierarchical Clustering (Academic Journal, 2010) Vol. 39 Issue 2, p234.

[2]. G. P.Dubey, N. Gupta, R. K Bhujade, A Novel Approach to Intrusion Detection System using Rough Set Theory and Incremental SVM (International Journal of Soft Computing and Engineering (IJSCE), 2011) 2231-2307, Volume-1, Issue-1

[3]. http://kdd.ics.uci.edu

[4]. http://nsl.cs.unb.ca

[5]. http://caida.rg

[6]. http://wand.net.nz/wits

[7]. http://ita.ee.lbl.gov/html/traces.html

[8]. J. Koshal, M. Bag, Cascading of C4.5 Decision Tree and Support Vector Machine for Rule Based Intrusion Detection System (Computer Network and Information Security, 2012) , 8, 8-20.

[9]. Z.A.Othman, E.E. Eljadi, Network Anomaly Detection Tools Based on Association Rules (IEEE, 2011) 978-1- 4577-0753-7.

[10].N. K. Hoong, P. K. Hoong, I. K.T. Tan, N.M Muthuvelu, L. C. Seng, Impact of Utilizing Forecasted Network Traffic for Data Transfers) IEEE, 211) 978-1- 4244-8830-8.

[11].N.C Anand,C.ScoglioS, B. Natarajan, GARCH NonLinear Time Series Model for Traffic Modeling and Prediction (IEEE, 2008) 978-1-4244-2065-0.

[12].S.M.A Burney, S. Raza, Monte Carlo Simulation and Prediction of internet load using conditional mean and conditional variance model (ICCS-IX, 2007).

[13].C.Guang,G.Jian,D.Wei, A Time series Decomposed Model of Network Traffic(Springer, 2005) 338-345.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Mobile Phone Controllable Smart Irrigation System

Dhammini U, Rajesh G, Rama Murthy V, Monika M K
Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The increase in technological innovation has made it possible to take full advantage of crop production. The need of a more proficient irrigation system is important for the maximization of crops and the most efficient means of achieving this is with the use of internet of things to control and monitor farms status. Smart irrigation is a very germane component of precision agriculture as it helps farmers to avoid water wastage and improve the quality of crop growth in their fields by reducing evapotranspiration, land runoff and advancing other old usage methods among others. This research aims to use a smart irrigation system to control water distributed to crops. The system does so by connecting the irrigation system and the user's Smartphone through the internet and automating the process of irrigation with computer programming and micro computing. This research was implemented using a microcomputer, internet driver, water pipes and C++ programming language was used with Virtuino for the design of the user interface. The system was tested and the result was 95.5% efficient.

Keywords: Agriculture, Internet Driver, Microcomputer, Smart Irrigation, Virtuino.

I. INTRODUCTION

Agriculture has a key role to play on the economy growth of the country. Water utilization in the society is on the high rise day by day which may lead to the problem of water scarcity. Currently, farmers are finding it difficult in the agriculture sector and the task of irrigating grassland is becoming quite complicated for the farmers due to lack of steadiness in their work and carelessness because oftentimes they put on the motor and then forget to switch it off leading to water wastage [3].

In [3], Irrigation is a systematic method of artificially providing water to the land or soil that is being cultivated. Conventionally in parched areas with no or little rainfall, there is need for water supply to thefields either by means of canals, hand pumps or tube wells. However, there is limitation to this methodbecause there might be an increase in workload of farm labour leading often to problem such as overirrigation or under-irrigation, and leaching of soil.

According to [2], Modern technological method of implementing irrigation exercise much more efficiently is known as Smart irrigation. It involves the use of sensors, micro-computing and remote communication by farmers to control the application of water to their farm land.

The internet is an interconnection of network enabled devices globally for the purpose of communication, transfer and sharing of data and so on [2]. According to

[9], the internet of things (IoT) is the network of devices such as vehicles and home appliances that include electronics, software, actuators and connectivity which allows this equipment to connect, interact and exchange data.

An IoT based irrigation method allows water allocation to crops soil easier, faster and much more proficient by eliminating the need of manual labourers, time wasting decision making and extraction of adequate data from the environment for better distribution of water, fertilizer and nutrients to the soil [1]. Human errors are reduced by replacing manual irrigation with automatic valves and systems (e.g. forgetting to turn off a valve after watering the field) and also maximized energy, time, and important resources. The installation and configuration of smart irrigation systems is, in general, fairly straightforward which serves as an advantage in terms of usage by the general public [5].

II. REVIEW OF RELATED WORKS

[4] developed a GSM Automated and SmartIrrigation System Using Android. The research was based on finding a way to easily control the temperature and humidity of a greenhouse based on the plants requirement which is difficult to do manually. The system was implemented by connecting two mobiles using GSM. The GSM module and microcontroller are connected using UART
(Universal asynchronous receiver / transmitter). When the moisture sensor senses the low moisture content of the soil, it gives a signal to the microcontroller. The microcontroller then gives a signal to the called mobile (which is kept in the auto answering mode). The called mobile activates the buzzer. Therefore, when calling mobile calls, that buzzer is heard indicating the valve needs to be open by pressing the button in the called function, the signal is given back to the microcontroller.

The microcontroller gives signal to the valves which causes it to get open. The water is given to the root of the plant drop by drop, and when the moisture content becomes sufficient, the sensor senses this and gives back the signal to the Applications microcontroller and the buzzer becomes off. Then by pressing the button in the calling function again, the valve is made turned off. The power supply needed by the controlling system is +5V. The information sharing using only SMS limits the system.

[7] also developed a GSM Based Irrigation System. This irrigation system was designed to use a simple GSM (Global system for mobile communication) device for relaying commands to and receiving information from the system. The interface and communication between user and the designed system is by the means of SMS on GSM network if the user is within the range of 10m of designed system.

The system sends the status of power supply through Bluetooth or SMS on GSM network to the user. The system checks the water flow from the pump. If electricity is present but no water supply is available, system sends information to the user through SMS on GSM network. Temperature sensors and humidity sensors which are installed in the field take the measurement of humidity level of soil and environmental temperature and send this information to the user. The user sends data in the form of SMS on GSM network to start or stop the irrigation according to received information.

The system uses an Arduino board which is connected to a GSM modem in which the transmitter of Arduinois connected to the receiver of the GSM modem and the receiver of Arduino is connected to the transmitter of GSM modem. The GSM modem works simply like a mobile phone, when we call on the number of GSM modem; it moves into the auto answering mode and can operate according to the coding which has been done in the microcontroller.

The drawback of the system is its lack of autonomy in scheduling water distribution and its geographical challenge of a range not more than 10m.

[6] developed a Smart Irrigation System for a greenhouse farmer to monitor the soil moisture and temperature of their greenhouse remotely from their android device. The temperature of the soil is detected using temperature sensor and moisture content in soil is detected using soil moisture sensor. Alert message is sent to owner's phone when the sensor values condition is satisfied. The system involves two modes: the automatic and manual mode. The user made decision based on the mode he/she is, if the mode is automatic then automatic system turns ON which makes the system to be ON. And if mode is manual then system turns ON using the input from user application. The User launched the APP and displaying the following option:- (a) Mode. (b) System turns ON/OFF. (C) Control hardware. The user takes corrective action on the particular situation.

The system runs on the Raspberry pi and android operating system. The system's drawback is that it can only work with an android operating systems platform.

III. SMART IRRIGATION SYSTEM DESIGN

The system architecture comprises different modules which include: Sensors, Micro-computing chip, Internet driver. The architecture is presented in Figure 1.

A. Sensor Module

YL-38 soil moisture sensor is a hygrometer that is usually used to detect the humidity of the soil. So, it is perfect to build an automatic watering system or to monitor the soil moisture of plants. The sensor is set up by two pieces: the electronic board (at the right), and the probe with two pads, that detects the water content (at the left). The sensor has a built-in potentiometer for sensitivity adjustment of the digital output (DO), a power LED and a digital output LED.

B. Micro-computing Chip Module

Arduino is the micro computing chip employed in this research. An Arduino is an Open Source electronic prototyping platform based on flexible easy to use hardware and software. Arduino boards are able to read inputs - light on a sensor, a finger on a button, or a Twitter message - and turn it into an output - activating a motor, turning on an LED, publishing something online. The board can be told what to do by sending a set of instructions to the microcontroller on the board. To do so, the Arduino programming language (based on Wiring) and the Arduino Software (IDE) based on Processing is used.

ESP8266 is the internet driver used in this research. The ESP8266 is the name of a micro controller designed by Espressif Systems. The ESP8266 itself is a self-contained WiFi networking solution offering as a bridge from existing micro controller to WiFi and is also capable of running selfcontained applications. This module comes with a built in USB connector and a rich assortment of pinouts. With a micro USB cable, you can connect Node MCU devkit to a laptop and flash it without any trouble, just like Arduino. It is also immediately breadboard friendly.

C. Virtuino Module

Virtuio is an android application for monitoring sensors or controlling electrical devices via Bluetooth, local Wi-Fi or the internet.

D. Application Module

For the communication between IoT devices and their servers an application programming interface (API) is essential. An API is a set of subroutine definitions,

communication protocols and tools for building software [8].

E. Water Pump Module

The water pump is the part of the system that delivers water to the soil. The motor drive is used to provide power to the water pump as well as interfacing the pump to the Arduino.

IV. HOW THE SYSTEM WORK?

The soil moisture sensor which has been placed in the soil reads the soil humidity and sends the readings to the Arduino microcomputer. Arduino then sends the information read to the API through the internet driver using mobile application (Smartphone). If the moisture content is less than 3.8ml a command is sent back to the pump to initiate or release water to the soil and if otherwise to stop the operation. The farmer also receives information on the whole process of irrigation through the Smartphone.
The proposed smart irrigation system is designed Step by step process will done as shown in to autonomously decide on the moment of the flow chart and explained in detail.
Activation and deactivation of electromagnetic valves releasing water on the agricultural field. The main control unit manages the process of Autonomous decision-making. Thus, in addition automated control of the irrigation process, the user can configure an autonomous operation of the system

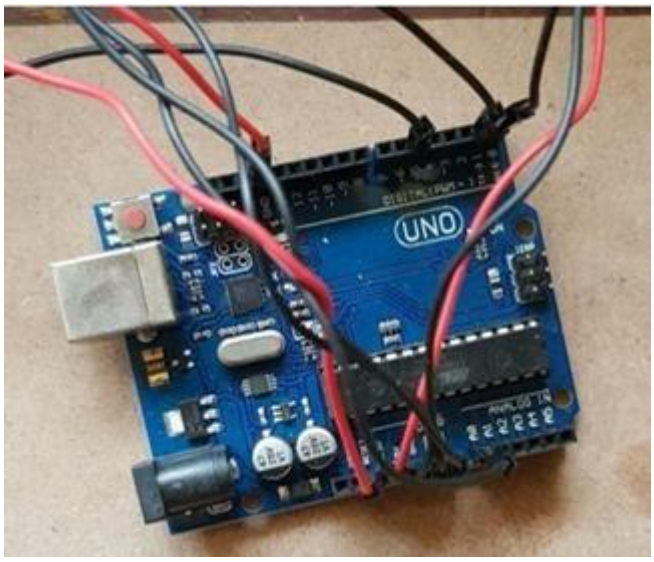


Figure 1: Arduino board with connections

ESP8266	Arduino UNO
VCC	3.3 V
CH_PD	
GND	GND
RX	Pin 4
TX	Pin 3

Figure 2: Arduino to ESP8266 connection

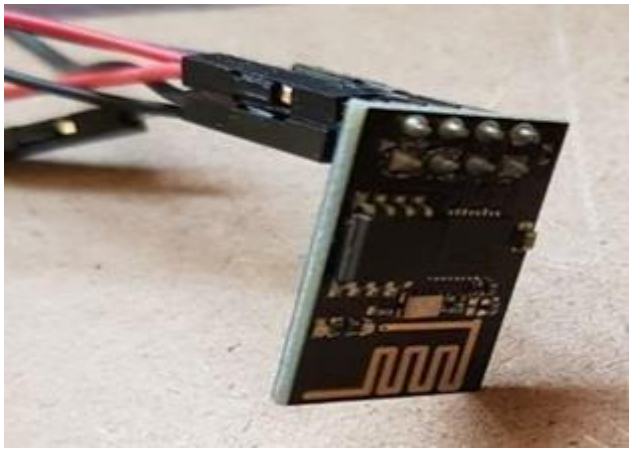


Figure 3: ESP8266

Figure 4: YL-38 Soil Moisture Sensor Module

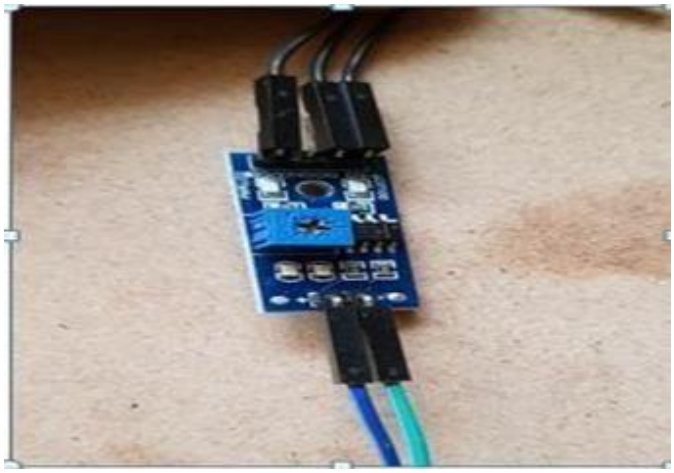


Figure 5: YL- soil moisture Probe

Figure 6: Water Pump

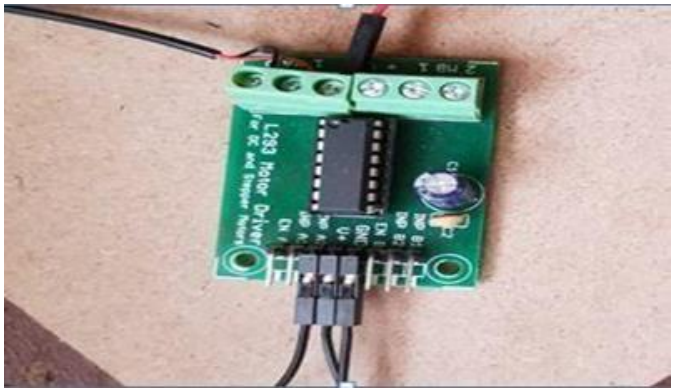


Figure 7: L293d Motor Drive

V. SYSTEM IMPLEMENTATION

Figure 2 presents the flow chart diagram of the designed system. Figure 3 to Figure 9 presents the hardware components in the implementation of the designed system. Figure 10 presents the sample test of the designed smart Irrigation System (a case study of a small flower pot).

The implementation of the designed system involves various modules and steps as described as follows:

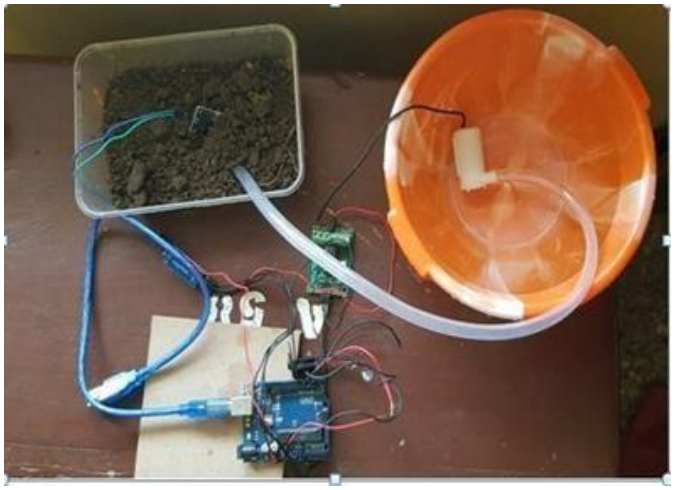


Figure 10: A Sample Test of the Designed Smart Irrigation System.

A. **Interfacing of the Arduino board with the Personal Computer** The following steps are taken to interface Arduino board with Personal Computer:

B. *Connection of the ESP8266 to the Arduino* The following steps are taken in connecting the ESP8266 to the Arduino: [1] Take a two pin connectors and cut them, strip them and connect them to one another in the connector. [2] Connect the two headed pin connector newly created to the VCC andCH_PD on the ESP8266 the other one headed side to the 3.3V pin on the Arduino board . [3] Connectthe GND pin on the ESP8266 to the GND pin on the Arduino.[4]Connect the RX pin on the ESP8266 to the pin4 on the Arduino.[5] Connect the TX pinon the ESP8266 to the pin3 on the Arduino .

C. *Connection of Soil Sensor* These are the steps taken to connect the soil sensor. Connect two pin connectors to the YL-38 Soil moisture sensor module and a YL- soil moisture Probe. (NOTE: Wire polarity not important).[2] Connect the A0 pin on the soil Sensor andSensor and connect to A0 on Arduino. [3]Connect the GND pin on the soil sensor and connect to GND on Arduino. [4] Connect the VCC pin onthe soil sensor and connect to 5V on sArduino.

D. *Connection of Water Pump to the system* The following steps are taken to connect the Water pump to the system:

1. Connect the negative wire on the pump to the MA1 slot on the motor drive .
2. Connect the positive wire on the pump to the MA2 slot on the motor drive.
3. Connect the –V pin on the motor to the GND pin on the Arduino.

[5] Connect the Enable A pin on the motor to the 5V pin on the Arduino.

E Connection of API

For the communication between IoT devices and their servers an application programming interface (API) is essential. An API is a set of subroutine definitions, communication protocols and tools for building software. For this project 'Thingspeak API' was used.Setting up a Thing speak channel

1. Signup to the thing speak website.
2. up to the thing speak website.
3. Fill new channels name.
4. Create fields for the recorded data, in this case two.
5. Go to the API key option. Generate read and write key.
6. Go to channels field and set result parameters to 5.

VI. RESULT

In conclusion the main purpose of this research is to develop a smart irrigation system with the functionalities to detect the soil moisture level, apply water to the soil if the moisture level is below a specified standard and interact with the user of the system through Smartphone. The implementation of the system involves a combination of hardware and software including: a micro-computing chip, a soil sensor, an internet driver, a water pump and a motor driver. The system was tested using a flower pot and the result was 95.5% efficient.

The Smart Irrigation System is meant to meet the need of a farmer with a small scaled plantation and can also be further extended to be useful for a large scaled agricultural plantation farmer.

VII. REFERENCES

[1]. P. Lahande,andDr. B. Mathpathi (2018), "IoT Based Smart Irrigation System," International Journal of Trend in Scientific Research and Development (IJTSRD), ISSN: 2456-6470, Vol 2 Issue 5, pp 359-362. https://doi.org/10.31142/ijtsrd15827

[2]. C. N. Lakshmiprasad, R. Aashish, and M. J. Syed "IoT Based Smart Irrigation System," Journal of Electrical and Electronics Engineering (IOSR-

JEEE), e-ISSN: 2278-1676, p-ISSN: 2320-3331, Volume 9, Issue 6, ver 1, pp34-40, Nov-Dec 2014, www.iosrjournals.org

[3]. G. R. Kumar, T. V. Gopal, V. Sridhar, G. Nagendra, "Smart Irrigation System," International Journal of Pure and Applied mathematics, Vol 119 Issue15, pp1155-1168.

[4]. H. N. Kamalaskar, and Dr. P. H. Zope, "Survey of Smart Irrigation System," International Journal of Engineering Sciences and Research Technology, ISSN 2277-9655, vol3 issue6: pp 224-228, June 2014.

[5]. J. Ravichandran, "Based on GSM Automated and Smart Irrigation System Using Android," International Journal of Trend in Scientific Research and Development (IJTSRD),

[6]. International Open Access Journal, ISSN No:7. 2456 – 6470 volume 3, Nov-Dec 2018https://doi.org/10.31142/ijtsrd18944

[7]. K. BhagyashreeandProf. J. G. Rana, "Smart Irrigation System Using Raspberry PI," International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Volume: 03 Issue: 05, August 2016

[8]. L. A. Vishal, P. B. Arti, A. Tejaswini, K. Amit, and Prof. P. SaiPrasad, "Smart Irrigation System," International Journal of Scientific Research in Science and Technology (IJSRST), Volume 2 Issue 5, Print ISSN: 2395-6011, pp343-345, September - October 2016.

[9]. "Drainage Engineering" by Luthin J N

[10]."Irrigation and Drainage Engineering" by Peter Waller and MulunehYitayew

[11]."Irrigation and Drainage: Sustainable Strategies and Systems" by PetreKozel

[12]."How to Make and Use the Treadle Irrigation Pump (Workshop Equipment Manual)" by Carl Bielenberg and HughAllen"Micro Irrigation Systems in India: Emergence, Status and Impacts (India Studies in Business and Economics)" by P KViswanathan and M Dinesh Kumar

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Ethical Hacking and The Need for Cyber Security

S Revanth Kumar, Sufiyan Pasha K, Ranganath, Amith S
Department of Computer Science & Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Hackingis one of the most dangerous diseases where the whole world is suffering. This the project focuses on the brutal attacks as well the consequences of crime committed in our society. Provides complete picture and preventive measures so to solve the problem Hacking problem. Today's generation is still lagging behind to solve the problem of hacking and hacking attacks preventive measures in solving this global problem growing day by day. To solve this problem of attack and robbery sophisticated security tool was developed.

That's why we should start thinking about hackers psychology as a primary means of preventing and stopping attacks by understanding their needs or desires. Establishment of the internet has solved many problems and brought many new things in this world like electricity trading, it's easy access to reference stores, in partnership computer, email, and new marketing methods as well the dissemination of information, but at the same time emerged in a very dangerous problem called robbery. State-owned enterprises, and private citizens around I the world longs to be a part of this rebellion, but they fearing that other hackers will invade their Web site server and change their logo with pornography, read their email, they stole their credit card number from line shopping, or install software that will do so privately transmit the secrets of their organization to the open internet. This study describes the ability, attitude and how this will be done to help customers find and connect security holes and the problem of moral fraud is explained once and for all with global problems and solutions to those problems listed.

Cyber Security accepts a major role in the field of information technology. Information security has become a major issue these days. Cybersecurity is a key element in the concept of 'cybercrime' which is rampant on a daily basis. Various governments and organizations are taking steps to curb cybercrime. Despite the different measures cybersecurity security remains a major concern for many. This paper places great emphasis on cyber security and cyber terrorism. Important cybersecurity trends and cybersecurity effect discussed online. Cyber-terrorism can cause organizations to lose billions of dollars in corporate space. This paper also describes the components of online terrorism and their incitement.

I. INTRODUCTION

Ethical hacking is an authorized practice by which a company can identify a potential data breach or loopholes which might cause data theft, that could cause serious consequences to a company.Ethical hackers use the same tools and methods as attackers, but they do not harm targeted programs or steal

information. Instead, they check the security of the system and report to the company the risks identified and instructions for resolving them.

Hackers and the malware that they use in crimes have seen great improvement and the methods they use are being highly sophisticated. Hacking can be classified in three different categories, depending on the intent and the motive of the Hacker. The types mainly being White, Black and Grey.

Fig 1. Types of ethical hackers

1. Black Hat Hackers:

These hackers are also popularly known as Unethical Hacker or a Security Cracker. These Hackers are knowledgeable but use their knowledge for wrong intentions such as stealing sensitive or personal data, compromising entire computer systems, or altering critical networks, stealing money/ransom for achieving their illegal goals.

Popular targets: Bank or companies with weak security.

2. White Hat Hackers:

These Hackers are popularly known as Ethical Hackers or Penetration Testers. These Hackers are Cybersecurity experts who use their skills to find vulnerabilities in organizational networks, computer systems and resolve such loop holes and protect the organization from such activities.

Purpose: Focuses more on security and protecting IT systems.

3. Gray Hat Hackers:

These Hackers are a hybrid of Black and White Hat Hackers. They are not certified like the White Hat Hackers, but their intention is for their own personal gain. Their purpose is not to hurt people nor want to help people. Rather they crack defences and find loopholes in systems.

Cybersecurity is a process of recovering and protecting computer systems, network devices, programs from a svarious Cyber-attack. With evolving technology, data is becoming more prone to attacks and breaches. Hence, to protect data for an individual or a customer we need new intelligent methods alongside strong pass codes to avoid unauthorized access and data leaks. Companies need this emerging technology to keep their data and systems away from unauthorized users accessing their network or data, which in return protects their users and also the organization's data.

II. ROLES AND RESPONSIBILITIES

Ethical hackers are expected to follow certain instructions in order to commit a formal robbery. An Ethical Hacker must abide by these rules and follow all ethical guidelines, out of which the most important rules of Ethical Hacking are:

1. They must obtain authorization and complete approval of the system which is owned by a particular organization.
2. Determine the range of their analysis and let the organization know the plan to solve the vulnerabilities.
3. In case of any security breach or vulnerabilities found, the organization must be informed about how it occurred in the system or network.
4. After finding a loophole or a vulnerability, keeping the documentations safe and confidential is important. A good hacker will always have a practice of wiping his traces after checking the system for any vulnerability. This will prevent malicious criminals from entering the target through the targeted holes.

Major advantages of Cybersecurity for businesses are:

1. Data protection from unauthorized access
2. Protection from deletion of data
3. Prevention from financial frauds
4. Improving customer's trust
5. Avert data theft or data leak

Major advantages of Cybersecurity for individuals are:

1. Reducing online shopping scams
2. Prevention of data falling into wrong hands
3. Protection from theft of passwords.
4. Help in building safe online transactions.
5. Secure social media identities.
6. Improves safety for kids and minors who lack knowledge of such online threats.

Implementing Cybersecurity by an individual:

1. Using strong passwords which is a combination of uppercase, lowercase, numeric and special symbols.
2. Not using the same password across multiple platforms or accounts.
3. Not sharing private information like bank details, passwords.
4. Using an antivirus protection on your system can help prevent malicious programs obtain access of your files or data.
5. Making use of closed source networks such as home Wi-Fi for banking purpose, which reduces the risk of data leakage.

III. CASE STUDIES

1. Cyber Security in E-Governance

E-Governance is an extension of the efforts completed by governments to restore relations with their nations. With its deep focus and acceptance, in terms of internet standards, E Governance extends to governments closer to their citizens. Existing and potential dangers in the circle of cybersecurity are among the most genuine difficulties of the 21st century. Security policies, practices, and techniques must be set up just as the use of security technology. It helps to secure e-Government systems against attacks, detects good management practices and has a proven course of suspension. An open private organization is an important part of cybersecurity in E Governance. These organizations can withstand coordination issues. Powerful activities to prevent cybercrime and to prosecute all appropriate ICT situations.

2. KasperskyKidnapping Case

The "highest" case of online surveillance, spying, and kidnapping included Ivan Kaspersky, the child director and CEO of Kaspersky Lab in Russia, who stands out among the world's most mysterious cybersecurity companies. Ivan Kaspersky was abducted for pay in 2011 while on his way to work from his residence in Moscow. As Russian media have shown, beginners — responsible veteran couples — conspire and register their child and two friends as a "muscle" of the plot. The hackers reportedly obtained all the necessary information from Kaspersky's client profile at Vkontakte, Russia's most popular social networking site administrative centre. Kaspersky was compelled to call his father to forward payment. Captors may use the same wireless to deliver food or have geolocation control enabled.

3. Uber Case study

Data breaches happen every day, in too many places, but the risk of data breach doesn't necessarily depend on the number, it may also depend on the risk and damage it causes the company's revenue and impact on the users or account holders, one of the biggest recent data breaches is Uber. The worst part of this

attack is how the Uber handled the issue, this is a lesson to most companies what not to do. In late 2016 just two hackers were able to steel the Users personal data with includes names, phone numbers and email addresses. They managed to steal the details of 600,000 driver's licenses and have gained access to the Uber GitHub account with a cloud-based third-party service with information obtained from GitHub. They have found a way to access Uber user data on AWS. Ubers paid those two hackers $100, 000 to permanently destroy the whole data they obtained and not letting the users or the regulators about stolen information.

IV. TYPES OF CYBER ATTACKS

1. Phishing

The crime of stealing sensitive information is a common form of robbery. All of our inboxes and messaging apps are filled with phishing scams. These are hidden messages as an organization (Amazon, Netflix, etc.) or someone you trust and, in most cases, will deceive you into clicking a link or opening a link.

2. Bait and Switch Attack

By using trusted marketing methods such as paid advertising on websites, attackers can deceive you into visiting hazardous or malicious sites. If websites sell advertising space, it can be purchased by corrupt attackers. The bonafide ad can be replaced by a 'bad' link that can be used to download malware, shut down your browser, or damage your systems. Alternatively, the ad may link to an official website, but it will be designed to redirect you to a malicious site.

3. Login Key

A key logger is a software that, when downloaded to your computer, records all the keys. The key recorder will record every keystroke on the keyboard, all usernames, password and credit card number, etc., which will display all your data and personal information.

4. Denial of Service (DoS \ DDoS) Attack.

Denial Attack of Service is a hacking technique designed to fill your web server with a host of applications until it overloads the web server leading to website crashes. To do this, hackers will use botnet or zombie computers with one function, filling your website with data requests.

5. ClickJacking Attacks

In this method, the hacker tricks you into clicking on something different from what you thought you were clicking. A click feature can be a button on a web page that, when clicked, does another job, allowing others to control the computer. The host website may be unaware of the existence of the click feature.

6. Fake W.A.P.

A hacker can use software to create a wireless access point (W.A.P.), which can connect to the 'official' public domain W.A.P. which you use. Once connected to W.A.P. fake, hacker can access your data. To deceive you, the criminal will give W.A.P. a real obvious name such as' T.F. Green Aiport for free WiFi.

7. Stealing Cookies

Cookies in your web browsers (Chrome, Safari, etc.) store personal data such as browsing history, username, and passwords on the various sites we access. The hackers will send an I.P. (data) packets pass through your computer, and may do so if the

website you are browsing does not have an SSL (Secure Socket Layer) certificate. Websites starting with HTTPS: // are secure, while sites starting with HTTP: // (no 'S') do not have SSL and are NOT considered secure.

8. Viruses and Trojans

Viruses or Trojan programs are malicious software that, once installed on your computer, sends your data to a cybercriminal. He can also lock your files, spread them across all computers connected to your network, and perform many other nasty acts.

V. PREVENTION TECHNIQUES AGAINST CYBER ATTACKS

The power to prevent online terrorism lies in the ability to securely secure cyberspace. Cybersecurity has an interesting similarity to terrorism. They are both different. Ensuring the security of information, data, and books is surprisingly difficult than breaking into a framework. The attacker has an unalterable preference for both common terrorism and cyber attacks. Due to government-sponsored attacks, the difficulties are even greater. Governments should ensure that their laws contaminate cybercrime and make them more realistic and adhere to them; it is important that biosphere countries take steps to ensure that their disciplinary and technical laws are sufficient to address the complexities posed by cybercrime.

The availability, confidentiality and integrity of information in any organization are essential efforts that must be made to ensure that they are incredibly secure because it is an important online resource that makes each organization stand and is currently dependent on it. The information has entered by the "cyber-terrorist" is something beyond records which may incorporate messages, web applications, web pages, and just as some indispensable operating systems.

Numerous methods to prevent attacks include:

1. Cybersecurity Hardware like VPN-capable firewall:

A firewall with a VPN will allow you to encrypt all communications whether you are in the office, home network, or on the Wi-Fi coffee shop. It will prevent Remote Desktop Protocol attacks (which we find to be the most common source of ransomware attacks), will infiltrate and block any intrusion attempts, and can serve as a filter to ensure that users cannot access potentially harmful websites while. logged in again. To save on costs, you may be able to access a used firewall online. If you purchase a used firewall, be sure to install the latest firmware and reset it before adding your team members to the system.

Make sure you change the default password, and make sure that any firewall you purchase has at least two verifications. On most firewalls, they will come with part of the app style for employees to install on their devices. This will extend the sign-in process by about 15 seconds each time but can greatly reduce your risk of breaking the data agreement.

2. Security keys:

This technique gains a lot of traction simply because of the number of successful brute-force attacks where attackers simply guess the right combination of letters, numbers and (hopefully) special characters. We recommend purchasing the key for each team member. They will then connect the key to their device, click the button, and with the appropriate configuration will gain access to the tools, documents, and programs they need. It is dangerous to think that none of your employees can be deceived by criminal attempts to steal sensitive information via email.

The most devastating internet attacks we see are the result of detailed social engineering campaigns that override standard verification without stopping any red flags. A complex attacker, similar to the one we must deal with on a daily basis, takes his time to learn style and tone and begins to send anonymous emails like regular correspondents, even using their email addresses. In some cases, the attacker may jump into the middle dialogue to switch back and forth to make you click on certain links or download attachments.

3. Settings and Configurations:

a) Multi-factor authentication

Two-factor authentication is a security feature available on most everyday devices and programs, yet many are reluctant to use it. However, we cannot stress enough the importance of this security provision. Here's how it works: When a user successfully logs in with his username and password, a text message or notification is sent to another device provided by the user, perhaps via his smartphone. . The user will then be asked to enter the code sent to their device, thus granting them access as usual.

Multiple Items authentication effectively prevents malicious players from accessing your email address and password information through malware, phishing scams or other methods. If malicious actors try to login using your credentials but cannot retrieve the code provided via text message or app, they will not be able to gain access to your system.

To keep an eye on this, you should set up an alert that will attempt to login from any malicious IP address where your username and password were entered, but hackers can recover the code sent to the user's device. The inability to log in prevented them from logging in.

b) Use Password Vault

With the total number of passwords, you need to manage these days, it is very common to fall into the trap of using the same password on multiple accounts. With Password Vault you only need to know the password to unlock Vault. The software will automatically fill out forms, websites, applications and system login with very complex and random passwords ***** it is almost impossible for an attacker to do so. A recent article in PC Magazine reviewed the ten most popular password managers.

c) Auto-update

We work with businesses that are constantly attacked by the internet; The common theme we see in all of these cases is that the operating systems of the victims and the software were outdated and without the latest protections. Operating system vendors regularly integrate their systems to tighten their security settings as new and detected threats are detected and detected. If you do not install your programs regularly, you are not protected by the latest security measures, which will put you at greater risk for cyber-attacks.

d) Malware scanner

Another low-cost security measure, Malware Scanner allows you to continuously detect malware threats. There are many malware scanners in the market in price and functionality. Like any security measure, it is important to make sure that your malware scanner's system and definitions are up-to-date—otherwise, the scanner will not be able to properly detect vulnerabilities.

e) Full-disk encryption

Every device in your business/home must be fully encrypted, this includes your mobile devices used for work, external hard drives, and even your NAS

devices. As more and more employees start working from home or away from the office, this is a low-cost way to make sure your business equipment is safe in case it ever gets stolen.

f) Lock your IP address

Whenever you are migrating your website host, lock down the IP address from which you administer the site. Then, enable two-factor authentication, which will give you three-factor authentication on top of effectively locking down your IP address.

If you are using a third-party tool like HubSpot or Pardot, they must have two-factor authentication and IP-enabled. This will prevent malicious actors from other regions or countries from getting into sensitive programs like your CRM.

g) Back up your system

If you put all your eggs in one basket, but the basket is burning, your eggs will probably not hatch. For the purpose of the quote, making a backup copy of your system is like placing your eggs in a non-combustible container and placing them in a basket. Make a backup copy of your systems regularly and store them on a fully encrypted device separate from your daily network.

Make sure your backup data is separated from the login credentials for your network. Backup clients for network-attached storage devices use separate and complex credentials so attackers can't encrypt your backups other than your workstations and servers.

Although more complex than many of the suggestions in this article, setting up a comprehensive information security program within your organization will expose you to your weaknesses in order to build appropriate security. While some information security programs can be handled by internal team members and a CISO, involving a third-party vendor to determine your risk profile can expose vulnerable configurations more effectively than those which are very closely associated with the system.

h) Continue to develop

One of the greatest pieces of advice we can give to businesses large and small is to always keep an eye out for new cyber security settings and potential threats. Other helpful resources to be monitored by the United States Computer Emergency Readiness Team, Symantec's annual Internet Security Threat Report, as well as software and software manufacturers for your business. You can also track the latest cybersecurity stories through stores such as Wired, Motherboard, and Info Security Magazine.

VI. CONCLUSION

"Cyber-terrorism" can in one way or another cause the death toll as it causes serious damage. While social media can use cybercrime, these organizations cannot stop using social media as they play an important role in organizational attention. Online terrorism has ensured many innocent lives and, meanwhile, provides many homes in a crisis situation that comes with psychological trauma to affected families. Cyber terrorism remains important in today's society. Not only is the war on Cyber Terrorism lagging behind, the current cyberattack attacks are becoming more and more powerful and tackling. Cybersecurity has an interesting similarity to terrorism. Ensuring the security of information, data, and books is surprisingly difficult than hacking into a system.

VII. FUTURE SCOPE

Over the next five years, cybercrime could wreak havoc on information technology. According to

researchers they estimated losses close to 6 trillion dollars. Therefore, it will be much brighter for people who work and resolve issues related to cybercrime and provide all the necessary security measures. Large corporations such as CISCO are fully integrated with network technology which is one of the top millions with millions of vacancies related to cybersecurity security because it is the future of information technology security. They are also a wide range of opportunities in government-related sectors and the security sector to save secure international data from cyber attackers.

VIII. REFERENCES

[1]. Jennifer Callen, Jason E. James, Cybersecurity engineering : The Growing Need
[2]. Ahmad Mtair AL Hawamleh, Alorfi, AlmuhannadSulaiman M, Alorfi, AlmuhannadSulaiman M, Ghada Al-Rawashdeh, Cyber Security and Ethical Hacking: The Importance of Protecting User Data
[3]. IEEE journals and proceeding papers
[4]. Ethical Hacking by C. C. Palmar ;IBM research division
[5]. Georg Thomas, Oliver Burmeister, Gregory Low, The Importance of Ethical Conduct by Penetration Testers in the Age of Breach Disclosure Laws.
[6]. Kanimozhi V.R and Shanmugapriya N, THE NEED FOR CYBER SECURITY
[7]. Vinitha K. P, Ethical Hacking
[8]. Prabhat Kumar Sahu, Biswamohan Acharya, A REVIEW PAPER ON ETHICAL HACKING
[9]. NEERAJ RATHORE, ETHICAL HACKING & SECURITY AGAINST CYBER CRIME
[10]. Ronald E. Pike, The “Ethics” of Teaching Ethical Hacking

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Smart Mirror with Voice Controlled Home Automation Using Raspberry Pi 4

Abhinav Gupta[1], Dr. Aparna K[2]

[1]Student, 6th Sem MCA, Department of MCA, BMS Institute of Technology, Bengaluru, Karnataka, India

[2]Associate Professor, Department of MCA, BMS Institute of Technology and Management, Bengaluru, Karnataka, India

ABSTRACT

This article introduces a design of the voice-activated interactive Smart Mirror. The mirror is transformed into a personalized digital device equipped with peripherals such as Raspberry PI4, Google Assistant, and a two-way mirror-covered LED Display that offers the most common conveniences like weathered., breaking news and local time. With the help of voice processing techniques, we can also control appliances or householding appliances such as lights, fans, and blubs, etc.

Keywords—Raspberry Pi-4, Google Assistant, Smart Mirror

I. INTRODUCTION

The Internet has transformed our lives. Mobile phone is now a smartphone, it gave rise to the concept of Internet of Things (IOT). Our research is to design a device called "Smart Mirror". This project was developed in an era where every day we see new technology coming out. The motivation of this project is to help people manage time, which has become an important aspect of life.

We know that the mirror's sole purpose is for self-grooming / self-admiration or to be used as a decoration, so we will design a prototype that combines both traditional mirror and time spent for smartphones by integrating various electronic features into the Smart mirror. The hands will be released. The rest of the document is organized as follows: In Part II some related work is presented. Section III presents the proposed work. Part IV deals with the problem definition. Section V of the algorithm.

II. RELATED WORK

Multi Display in Black Mirror [1] through Toshiba is a prototype which mixes the features of pill collectively with reflecting floor of mirror. Taking into consideration one of a kind domestic environment it gives one-of-a-kind configurations: washroom and kitchen. Consider the washroom whose prototype gives records beneficial for the start of day which includes the climate forecast and health records from private gadgets while in kitchen, even as getting ready recipes it permits the consumer to engage via gestures for equipment manage because it has an optical camera.

"Smart Mirror: A Reflective Interface to Maximize Productivity"[2] Abhinav Gupta designed a prototype

which describes a voice-managed wall mount reflect able to showing multimedia contents and additionally facial recognition. It makes use of OpenCV (Open-Source Computer Vision Library) as an open-supply laptop imaginative and prescient and gadget getting to know software program library.

DESIGN AND DEVELOPMENT OF INTELLIGENT MIRROR USING RASPBERRY PI"[4] through the medium of Abhinav Gupta designed a futuristic interactive mirror multimedia using artificial intelligence for home environment and even for commercial use in factories can display weather, breaking news and time.

"IMPLEMENTATING OF MAGIC MIRROR OS USING RASPBERRY PI 4 model b+ via way of means of Abhinav Gupta designed and carried out a voice-managed Jarvis named Assistant wall replicate Mirror, called “Smart Mirror”.

It is a tool which can characteristic each as a replicate and an interactive show showing multimedia content material as time, date, climate and information simultaneously.

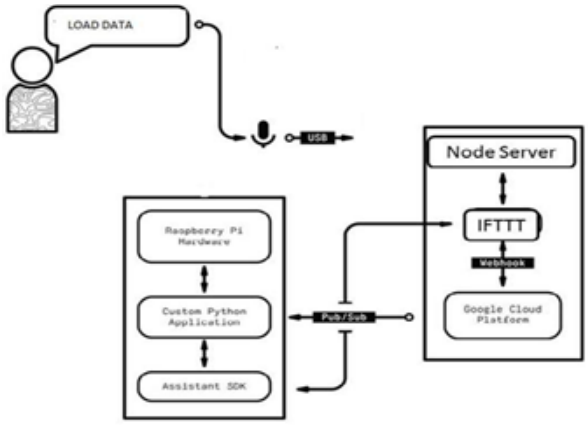


Fig.1-Block diagram

And the display is located in the back of unmarried sided reflect so the display is off at that motion the display seems as reflect and while we begin the reflect it'll display statistics will show. The controller can even connect to the voice enter i.e., google assistant. When we command to google assistant then in keeping with command the statistics will fetch through controller from server and show on display. In our undertaking we're going to show the present-day time date and surrounding temperature and news.

III. ALGORITHM

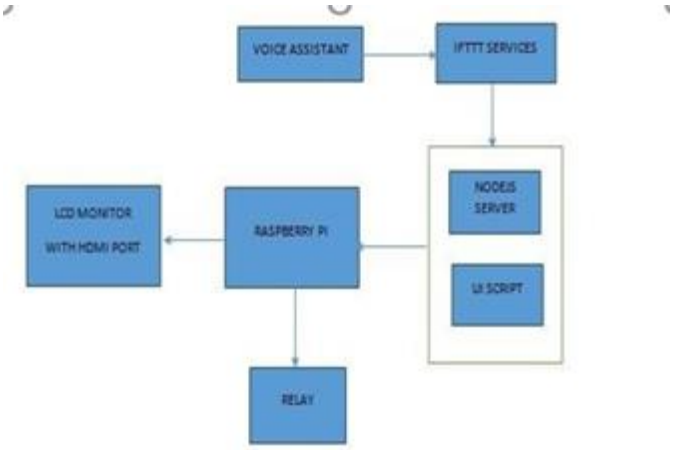
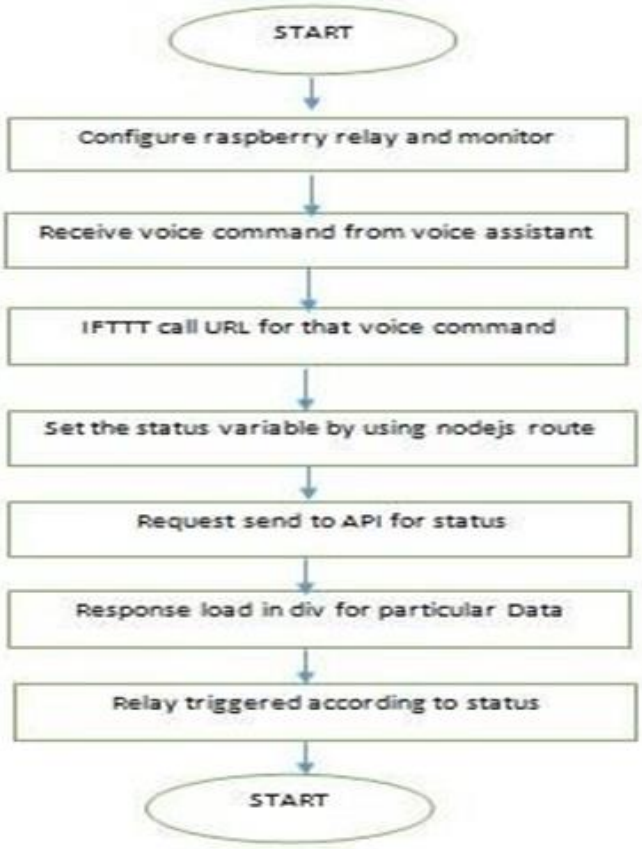


Fig.2 - Block flow of proposed model.



IV. PROBLEM STATEMENT

Along with counsel, humans significantly cost their appearance, spending tremendous quantity of time in the front of the replicate in the course of their morning and night time routines. This is a considerable quantity of time spent wherein crucial matters are taking place, however the thoughts aren't working. As our lives are more and more spent looking at phones, laptops and televisions. It also can purpose cataracts, headaches, eye stress and discomfort. Digital eye to as pc imaginative and prescient syndrome, is stress, on occasion referred nicely documented. If you've ever skilled dry or irritable eyes, blurred imaginative and prescient, eye fatigue, or head, neck and again ache after the usage of a pc or Smartphone, then you're acquainted with it. A compact product is needed to allow someone to do well whatever they want to get ready for the new day with voice control that really doesn't cause eye strain and also turns the device on or off Computerized household appliances in the home such as lights, fans, blubs, etc.

STEP 1: Start the power supply of raspberry and monitor by connecting each other with HDMI convertor.

STEP 2: Configure Monitor and relay Module 4x3 with raspberry pi with jumper wires

STEP 3: Load Bash script for auto run files.

STEP 4: Load the Web service in Kiosk mode on start up.

STEP 5: Load The web service in KIOSK mode.

STEP 6: The web service will load data on screen on request of voice assistant.

STEP 7: Voice command from the assistant generated it will invoke web service in backend for particular command using web hook by IFTTT.

STEP 8: That service will set the status variable for particular field like temperature, news;

STEP 9: The status will be read by JavaScript and call corresponding API for data.

STEP 10: Response got from API will be loaded on front end web service.

STEP 11: Load Python GPIO script for relay Module operation in start-up.

STEP 12: Uniformly for voice command python script will read the response by the system.

STEP 13: The data and Command received on python Scripts will be process to change state of relay Module.

STEP 14: Using the state of relay NC NO we can control appliance.

Document proposes a smart mirroring system that allows users to use an object as an interactive interface to provide the 's customizable services. Health monitoring is a boon. Supplement for a healthier life. With the feature of controlling the lighting settings, it can be applied in many different fields such as beauty salons, hotels, in addition, for security purposes, facial recognition and recognition techniques can be used. personalized voice. This ensures that only authenticated users can access the mirroring information.

V. REFERENCES

[1]. Toshiba: Toshiba to Unveil Leading –edge Technologies at CES 2014.

[2]. Piyush Maheshwari, Maninder Jeet Kaur0, SarthakAnand, "Smart Mirror: A Reflective Interface to Maximize Productivity", International Journal of Computer Applications (0975 – 8887), Volume 166 – No.9, May 2017

[3]. Vaibhav Khanna, Yash Vardhan, Dhruv Nair, Preethi Pannu," DESIGN AND DEVELOPMENT

OF A SMART MIRROR USING RASPBERRY PI “, International Journal of Electrical, Electronics And Data Communication, ISSN: 2320-2084 Volume-5, Issue-1, Jan.-2017

[4]. Suryansh Chandel, AshayMandwarya, S.Ushasukhanya,"IMPLEMENTATION OF MAGIC MIRROR USING RASPBERRY PI 3”, International Journal of Pure and Applied Mathematics, Volume 118 ,No. 22 2018, 451-455 ISSN: 1314-3395 (on-line version) url: http://acadpubl.eu/hub, Special Issue.

4th National Conference on Advancements in Computer Science & Engineering
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

A Research Paper on Wireless Sensor Networks

B Abhishek[1], C Harshavardhan[1], D Jaswanth Reddy[1], E Sai Jaswanth[1]
[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

In recent times, it has been witnessed that wireless systems based on IoT-based have developed rapidly in various sectors. The IoT (Internet of Things) is the network in which physical devices, equipment, sensors and other objects can communicate among themselves without human involvement. The WSN (Wireless Sensor Network) is a central component of the IoT, which has proliferated into several different applications in real-time. The IoT and WSNs now have various critical and non-critical applications impacting nearly every area of our everyday life.

Wireless Sensor Networks (WSNs) play a major role in revolutionizing the world by its sensing technology. WSNs has emerged as that powerful technology which has multiple applications such as such as military operations, surveillance system, Intelligent Transport Systems (ITS) etc. WSNs comprises of various sensor nodes, which captures the data from the surrounding alongside monitoring the external environment. Much of the research work is focused on making the sensor network operating with minimum consumption of energy, so that it can survive for longer duration. The primary concern in the direction of saving energy has been due to the discharging of those batteries on which sensor nodes are operated. In addition to that, WSNs are also exploited for its security aspects so that it can be used in some confidential sectors like military battlefield. This paper, introduces the WSN in different aspects like applications, routing and data collection, security aspects and also briefs about simulation platform that can be used in WSNs. This paper contributes in a fashion about introducing the WSNs in different sectors of its operation and reflecting its significance.

Keywords: Introduction to WSN, Routing, Simulation platform in WSN, security aspects in WSN, applications of WSN.

I. INTRODUCTION

A Wireless sensor association can be portrayed as an association of contraptions that can pass the information gathered from a noticed field on through distant associations. The data is sent through different centers, and with an entry, the data is related with various associations like distant Ethernet.

Headway in remote correspondence has made conceivable the advancement of remote sensor networks involving gadgets called sensor hubs. Sensor hubs are low power, little size and modest devices,capable of detecting, remote correspondence and calculation. When the sensors are conveyed in the organization they arrange themselves and interface with one another for information assortment

and along these lines sending the information to the Base Station.

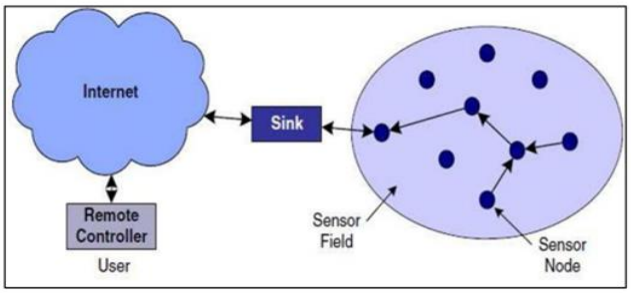


Fig.1 Architecture of a typical WSN

WSN can likewise be characterized as an organization including conceivably low-size and low-intricacy gadgets named as hubs which are fit for detecting the climate and imparting accumulated data from the checked region; the assembled information can be communicated straightforwardly or through multi - bounces to sink, which can then utilize it locally or is associated with different organizations (for example web) through gateway hubs.

The fundamental parts of sensor hub comprise of a detecting unit, a handling unit, a handset and a power unit as displayed in the Figure 2. Detecting unit detects the actual amount which is then changed into advanced one through ADC for example Simple to Digital converter. From there on processor is utilized for additional calculations and handset is utilized to communicate and get information from different hubs or from the Base Station.

Power unit is the most unmistakable unit in any sensor hub. When the battery is depleted, it can't be traded for unattended applications. Different units are application subordinate unit like Mobilizer, Power Generator and Location FindingSystem.

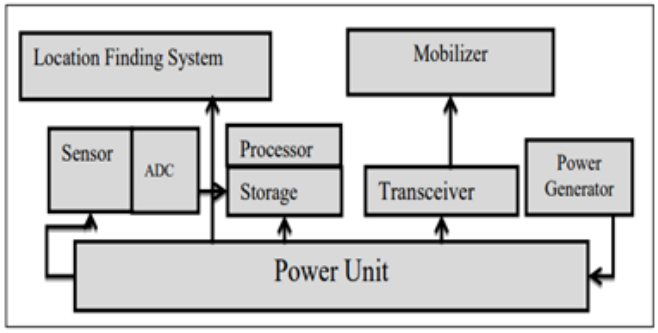


Fig2. Components of a sensor node

II. ROLE OF IOT IN WSN

Huge grouping suppositions and overviews of WSN and IoT-based energy-saving advances have been upheld by a few exploration papers and studies. All through this segment, a portion of these significant artistic works are audited, which present their principal regions and various classes distinguished by them.

The paper introduced the plan as well as the achievement of sun oriented energy controlled accuracy agrarian (PA) network with the WSN by using IoT design to satisfy the prerequisite of distinguishing incredibly compelling ways for a brilliant farming administration framework. This gave framework gave ranchers valuable data in an easy to use and simple to get to way with constant information correspondences through IoT about saltwater interruptions, the dampness of soil, level of water, wet circumstances, temperature and the general condition of the land. The creators in gave an investigation of IOT information gathering and the ideas of settling on a choice. The functional and upkeep overview of PV frameworks and WSNs in light of IoT for the observing of PV boards was introduced in.

The exploration recommended a methodology used to enhnance energy utilization in WSN-IoT ecological activities through the Chaotic Whale Optimization

Process. The aftereffects of energy proficiency comparative with other customary methodologies were gotten. The outcomes showed that in the WSN-IoT incorporated framework, the proposed approach accomplishes better energy productivity. The study was acted in on the deferrals, energies, butterflies, throughput, parcel conveyance proportions (PDR) from the perspective of WSN and execution of steering conventions was estimated utilizing latencies, transmission capacity, jitter and postponement. A calculation was intended to further develop AODV directing in IoT. Two tables were converged into one table, for example table of directing and web access table for convention advancement. This paper pointed chiefly to break down recreation investigations of the IoT AODV directing convention, and to use the NS2 test system to further develop AODV execution and IoT AODV execution. The most recent form is accessible. Likewise, WSN-helped IoT has numerous constraints, making it inconceivable for conventional steering conventions to be utilized straightforwardly. Energy is significant limitation for IoT gadgets helped by WSN. To impart among sensor hubs, more power is consumed than detecting and registering. Thusly, compelling energy the executives approaches are fundamental to expand the organization's life. In paper , the creator proposed an energy-cognizant multi-client and Multi-Hop Hierarchical Routing Protocol (EAMMH-RP) which covers Communication with Multi-Hop wherein energy is appropriated similarly across bunch development sensor hubs, a clever grouping of calculations for group variation and pivoting and an original energy utilization decrease component for long-range interchanges.

Sensors can be utilized to follow the environment and return the data for longer. A convention was proposed in which envelop a vigorous directing convention for IoT detecting organization. From the beginning, in the focal point of the organization field, a meeting region was constructed. The systems of grouping and multipath were used as it limits energy utilization and further develops unwavering quality. In the Castalia test system, the presented convention was mimicked to accomplish proficiency under various circumstances, like parcel transmission, normal energy use, start to finish postponements and organization life span. The directing calculations and models were evaluated in regarding progression boundaries, such as diminishing postponement, energy use and enhancing the information conveyance proportion. The IoT and WSN calculations in light of IoT were partitioned into two classes for grouping: energy cognizance, delay, throughput, information transmission and bundle misfortune mindful. The article upgraded the customary steering convention and presented an inventive convention with attributes like another information transmission framework and an improved strategy for determination .Consequently the hole of the WSNs in genuine world and the genuine heterogeneous setting was connected. With the assistance of execution estimations, the result of reproduction uncovered the difference between existing Hy-IoT and projected convention.

III. CHALLENGES OF WSN IN IOT

Different heterogenous ancient rarities introduced and conveying in various settings achieve IoT 's intricacy and make arrangement of safety components considerably more confounded. Existing WSN security research offers principally answers for emotional issues, without thinking about the effect of the IoT standards and highlights as analyzed in this archive.

a. Real time management

For asset controlled sensor organizations, it is a troublesome issue. All things considered, a proficient assistance passage configuration is required in the IoT framework to limit how much information to be sent by continually inspecting client information, and savvy information driven center product plan to impart ongoing data just while perusing more than edge.

b. Security and privacy

In genuine applications, wellbeing, trust and protection are additionally significant issues. The method for accomplishing various degrees of security is both troublesome and delicate. These wellbeing strategies are reasonable for M2 M organizations where the gadget and the server have a current trust relationship. Other than its typical sensor usefulness sensor hubs with this "IP to the field" worldview have extra liabilities. The sensor hubs will subsequently go up against new assignments or difficulties with this extra obligation. Three potential undertakings will be talked about: security, administration quality (QoS) and network design. Coming up next are tended to.

c. Quality of service

Concerning knowledge proposed to the sensor hubs, all heterogeneous gadgets of the web of things need to add to the nature of administration. This heterogeneous gadgets permit a conveyance of responsibility between the hubs with the assets available. The ongoing QoS approaches accessible on the Internet actually requires enhanecemnt because of dynamic organization designs and connection highlights.

d. Configuration

Alongside QoS the executives and security, sensor hubs need to oversee different undertakings, for example, organizing for the new hub joining the organization and ensuring self-mending by recognizing and erasing of imperfect hubs and tending to the board for developments of versatile organization and so forth. Notwithstanding, it's anything but a standard capacity of self-designing the most recent hub on the Internet. In this manner, the client should introduce the appropriate software and take sufficient measures to prevent device failures if this network setup is to run easily.

e. Availability

WSNs can be profited by presence of compromised hubs. To integrate encryption calculation for WSN security, additional expense ould be charged. Be that as it may, esearchers have created huge strategies in which some changed the code and reused it, a utilized strengthening interchanges to meet the objectives. Other than this, techniques have been intended to get to the information. Subsequently, need of accessibility is basic t protect the functional administrations of Wsn's. It likewise aid the support of the whole organization till its end.

f. Data integrity

WSN can be compromised when malignant hub enters the organization and infuses some unacceptable information or swaying remote channel debases the first information . For instance, if a maovelent hub moves the bogus information to the parcels got by the BS, it will influence the trustworthiness of information. however, the information misfortune or modification in information may be caused because of broken network. Along these lines, it is expected that information uprightness should be kept up with all through the transmission of information bundles.

g. Confidentiality

Security in IoT involves different difficulties, among which confidentialityis the significant perspective. The information is kept classified by picking encryption capacities like normal and shared secret key encryption calculations, e.g., the Blowfish, AES block code, and Triple DES. In any case, encryption process isn't adequate to safeguard protection of the information and data alone as a security system. A traffic investigation for the code information can be done by the aggressor so touchy information can be really distributed. Moreover, the malevolent hub can actually think twice about scope of other sensor hubs by utilizing a common gathering keypad and afterward awaken and disentangle delicate data.

IV. DATA AGGREGATION

As portrayed — WSNs are fundamental IoT blocks that have multiplied in a few different applications continuously. WSN hubs are normally little and battery-driven apparatuses. Along these lines, the life span of the organization is an essential thought for WSN information collection. During the assortment of information, various issues like expanded energy use, for example energy inadequacy and expanded life expectancy, were found.

Information total methodologies are generally used to safeguard satisfactory adjusting productivity in the appropriation of detected information. The reason for information assortment program is to really detain and appropriate information parcels that energy utilization, gridlock and organization life, information consistency, and so forth can be limited.

V. CONCLUSION

Headways in PC innovation have added to the development of WSNs, which whenever sense the essential boundaries. The IoT based WSN frameworks are acquiring colossal consideration lately. Regardless, during highlight point transmission, these frameworks experience the ill effects of confined data transfer capacity, power and assets. Information gathering is a celebrated technique for easing this issue.

A critical issue in sensor networks is the way significant data can be handled in a more energy-saving way. Consequently, different information conglomeration calculations were accustomed to decreasing the power utilization which is audited in this paper. In this paper, the current works characterizing the job of IoT in WSN is assessed and afterward the different information collection approaches proposed in past works is introduced. The information conglomeration strategies center around the energy preservation, lifetime improvement, better QoS and significant level security of the organization

VI.REFERENCES

[1]. Y. Cho, M. Kim, S. Woo, Energy Efficient IoT based on Wireless Sensor Networks for Healthcare, Int. Conf. Adv. Commun. Technol. (ICACT) (2018). [2] H.W. Kim, D. Kyue, Technology and Security of IoT, J. Korea Instit. Informat. Secur. Cryptol. 22 (1) (2012) 7–13.

[2]. S.A.H. Antar, N.M. Abdul-Qaw, S. Almurisi, S. Tadisetty, Classification of Energy Saving Techniques for IoT-based Heterogeneous Wireless Nodes, Procedia Comput. Sci. 171 (2020) 2590–2599.

[3]. N. Kaur, S.K. Sood, An Energy-Efficient Architecture for the Internet of Things (IoT), IEEE Syst. J. 11 (2) (2017) 796–805.

[4]. M. Healy, T. Newe, E. Lewis, Wireless Sensor Node hardware: A review, in: 2008 IEEE Sensors, 621-624, 2008.

TECHNO SCIENCE ACADEMY
THE INTERNATIONAL OPEN ACCESS PUBLISHER

IJSRSET

Fourth National Conference on Advancements in Computer Science and Engineering

NEW HORIZON COLLEGE OF ENGINEERING
Autonomous College Permanently Affiliated to VTU, Approved by AICTE & UGC
Accredited by NAAC with 'A' Grade, Accredited by NBA

Organised by
Department of Computer Science and Engineering, New Horizon College of Engineering
In Association with
VTU, Belagavi, Silver Jubilee Celebrations and Computer Society of India

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com
Email: info@technoscienceacademy.com